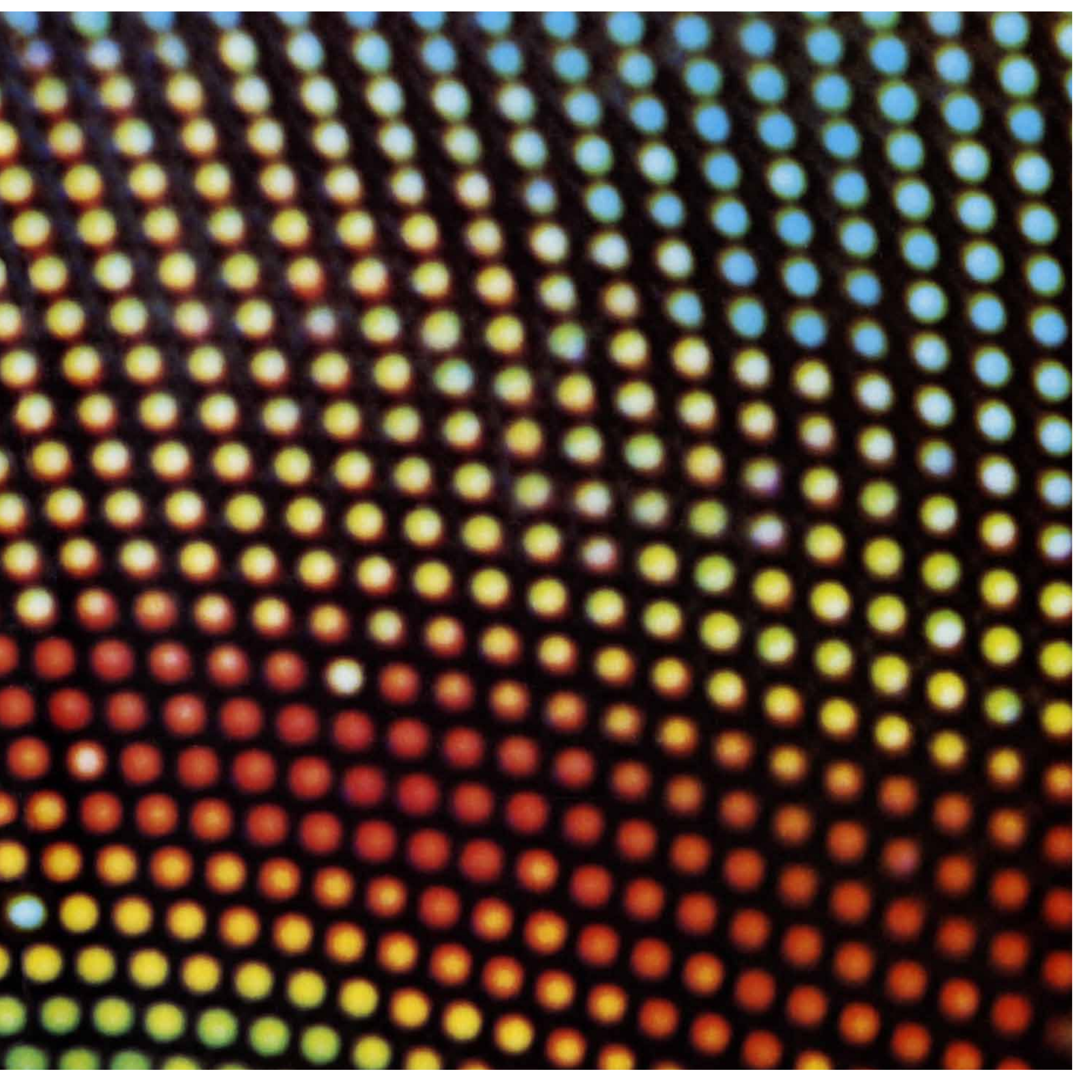
INVESTIGACION Y CIENCIA

Edición en español de **SCIENTIFIC AMERICAN**



OJO DE INSECTO

Septiembre 1977

140 PTAS.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 12
Septiembre 1977

INVESTIGACION Y
CIENCIA
Edición en español de SCIENTIFIC AMERICAN

EDITA Prensa Científica, S.A.
Calabria, 235-239
Barcelona-15
ESPAÑA

Colaboradores de este número:

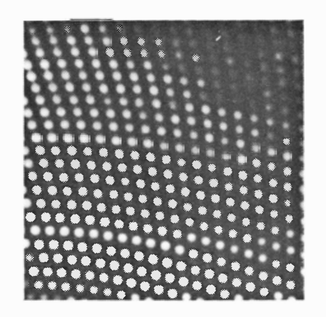
Asesoramiento y traducción:

Lluïsa Vilageliu: *Debate sobre el ADN recombinante;* Manuel Puigcerver: *La atmósfera de Marte;* José Luis Rodríguez Sánchez: *Hepatitis vírica;* Rafael Llinás: *La solidificación del cemento;* Nuria Ferrer: *Aniones de los metales alcalinos;* Olaguer Escolá: *El ojo compuesto de los insectos;* Luis Bou: *Gauss; Juegos matemáticos;* Manuel Puigcerver: *Taller y laboratorio;* Gerardo Espinosa: *Ciencia y sociedad.*



Libros:

Philip Morrison y
Luis Alonso



LA PORTADA

La fotografía de la portada es una toma extremadamente cercana de los distintos elementos, u omatidios, que integran el ojo compuesto de un insecto (véase "El ojo compuesto de los insectos", por G. Adrian Horridge), en este caso un tábano, *Tabanus lineola*, hembra. Los colores no son los propios naturales de los omatidios; sino que están originados por interferencias ópticas constructivas y destructivas.

Suscripciones:

Prensa Científica, S.A.
Departamento de suscripciones
Calabria, 235-239
Barcelona-15 (España)

Distribución para España:

Distribuciones de Enlace, S.A.
Ausias March, 49
Barcelona-10 (España)

Distribución para los restantes países:

Editorial Labor, S.A.
Calabria, 235-239
Barcelona-15 (España)

Precio de venta de este número:

España: 140 pesetas
Extranjero: 2 U.S. $

Condiciones de suscripción:

España:
Un año (12 números): 1.540 pesetas
Extranjero:
Un año (12 números): 30 U.S. $

Ejemplar atrasado ordinario:
160 pesetas

Ejemplar atrasado extraordinario:
225 pesetas.

Solicitado control **OJD**

PROCEDENCIA DE LAS ILUSTRACIONES

Fotografía de la portada de Fritz Goro

Página	Fuente
7	George V. Kelvin
8	Jack D. Griffith, Stanford University School of Medicine y Andrew Staehelin, Universidad de Colorado *(izquierda);* Jack D. Griffith, Stanford University School of Medicine *(derecha)*
9	© David Scharf *(arriba);* Jack D. Griffith, Stanford University School of Medicine *(abajo)*
10	Maria Schnoss, Stanford University School of Medicine
11	Jack D. Griffith, Stanford University School of Medicine
12-18	George V. Kelvin
21-22	Jet Propulsion Laboratory, California Institute of Technology
23-26	Jerome Kuhl
27	Jet Propulsion Laboratory, California Institute of Technology
28-29	Jerome Kuhl
30	Jet Propulsion Laboratory, California Institute of Technology
33	Guy A. Cabral, Baylor College of Medicine *(izquierda);* Clifton R. Gravelle, Center for Disease Control *(derecha, abajo)*
34	Gordon R. Dreesman, Baylor College of Medicine
35-39	Ilil Arbel
40	Ferenc Gyorkey, Veterans Administration Hospital, Houston
42-43	Ilil Arbel
49-50	Joandomènec Ros
51-52	J. D. Ros y Estudi Antoni Sellés
53	Joandomènec Ros
54-56	Joandomènec Ros
57-59	J. D. Ros y Estudi Antoni Sellés
63	D. D. Double y A. Hellawell, Universidad de Oxford
64	D. D. Double y A. Hellawell, Universidad de Oxford *(arriba);* Ilil Arbel *(abajo).*
65	D. D. Double y A. Hellawell, Universidad de Oxford
66	Ilil Arbel
67-68	D. D. Double y A. Hellawell, Universidad de Oxford
71	Ralph Morse
72-79	Alan D. Iselin
82-83	Tom Prentiss
84	G. Adrian Horridge, Universidad de Cambridge
85-87	Tom Prentiss
88	Type Systems Inc.
89-91	Tom Prentiss
92-93	Type Systems Inc.
96	Gabor Kiss
98	Burndy Library
99-102	Gabor Kiss
103	Dan Todd
104-106	Gabor Kiss
109	Ilil Arbel
110	Henry Regnery Company
111-113	Ilil Arbel
114	Fritz Goro
115-118	Michael Goodman
121	*The Magic Mirror of M. C. Escher.* © 1976, Random House

Dep. legal: B. 38.999-76
Fotocomposición: Fort, S.A.
Rosellón, 33 - Barcelona-15
Cayfosa. Santa Perpetua de Moguda.
Barcelona.

Printed in Spain - Impreso en España

Cartas

Señor Director:

En el interesante artículo "Semiconductores amorfos", de David Adler (INVESTIGACION Y CIENCIA, julio) se dice que hacia 1958 estaban en marcha estudios sobre esta materia que llevaron a la creación de dispositivos de gran éxito, como los diseñados por la Xerox Corporation sobre fotoconductores electrofotográficos.

Varios años antes de esa fecha, el tubo más corriente para cámaras de televisión, el Vidicón, ya había empezado a comercializarse. Dicho tubo se funda en una capa retinal sensible, construida deliberadamente de trisulfuro de antimonio, material fotoconductor. Este dispositivo, hecho de Sb_2S_3 amorfo, sigue produciéndose un cuarto de siglo más tarde.

STANLEY V. FORGUE

Ocala, Florida.

Señor Director:

La reseña de los libros recientes sobre virus lentos (INVESTIGACION Y CIENCIA, agosto) fue esclarecedora. Sin embargo, en ella se vierte una afirmación discutible. Sostiene el autor: "Entre los miembros del pueblo Fore de Nueva Guinea, las mujeres y los niños de ciertos grupos sufren de 'Kuru', enfermedad parecida, que va asociada con el rito extraordinario y recientemente adoptado de frotar con los sesos de un hombre que ha muerto a los parientes en duelo de la rama femenina de la familia". Las investigaciones realizadas por D. Carleton Gajdusek y otros sugieren que el Kuru está relacionado con un virus transmitido por canibalismo.

JOHN A. AGNEW

Syracuse University
Syracuse, New York

Señor Director:

En el artículo "Teoría del arco iris" de H. Moyses Nussenzveig (INVESTIGACION Y CIENCIA, junio), el autor afirma "La dependencia de los arcos supernumerarios con respecto al tamaño de las gotas explica por qué son más fáciles de ver cerca de la cima del arco iris: las gotas de lluvia tienden a hacerse mayores a medida que caen". Si esto fuese cierto, los supernumerarios se acercarían gradualmente al arco principal hacia la base del mismo. No obstante, los primeros supernumerarios mantienen un espaciamiento aproximadamente constante antes de desvanecerse cerca de la base del arco. ¿Cuál es entonces la explicación de la desigual distribución de supernumerarios en torno al arco?

La explicación reside en la distorsión aerodinámica respecto de la forma esférica en las gotas grandes en caída. Las gotas cuyo radio supere los 0,25-0,5 milímetros se van aplanando cada vez más en su parte inferior. Por tanto, no contribuyen a las partes horizontales (cimeras) del arco, porque un corte vertical de las mismas no es circular. Contribuyen a las partes verticales (inferiores) del arco, porque su sección horizontal sigue siendo circular.

ALISTAIR B. FRASER

Pennsylvania State University
University Park, Pennsylvania

Señor Director:

Respecto al artículo "Fobos y Deimos", de Joseph Veverka (INVESTIGACION y CIENCIA, abril), entiendo que Jonathan Swift hizo algo más que repetir la hipótesis matemática de Kepler acerca de la existencia de dos satélites marcianos. Afirmó que sus períodos orbitales eran de 10 horas y 21,5 horas, y que distaban tres y cinco diámetros de Marte respecto del centro de dicho planeta.

¿Tuvo otra fuente de información aparte de Kepler y de su despierta imaginación? Las probabilidades en contra del hecho de que alguien, antes de Asaph Hall, tuviese un telescopio lo bastante bueno como para observar esas dos lunas son formidables, pero quizás resulten aceptables si se las compara con las posibilidades en contra del hecho de que Swift supusiera con bastante exactitud el período de rotación y la distancia a Marte de los dos satélites.

BENJAMIN R. MOSER

Stanley, Virginia

Los autores

CLIFFORD GROBSTEIN ("Debate sobre el ADN recombinante") es catedrático de biología y vicerrector de relaciones universitarias de la Universidad de California en San Diego. Se doctoró en zoología por la Universidad de California en Los Angeles y luego trabajó como profesor ayudante de zoología en el Oregon State College. En 1946, después de haber trabajado tres años de fisiólogo en las Fuerzas Aéreas, ingresó en el National Cancer Institute como investigador, durante 10 años. Enseñó luego biología en la Universidad de Stanford durante otros nueve años. Entró en la facultad de San Diego en 1965, y desde 1967 a 1973 fue decano de la Facultad de Medicina y vicerrector del area médica.

CONWAY B. LEOVY ("La atmósfera de Marte") es profesor de ciencias atmosféricas y geofísica y profesor adjunto de astronomía en la Universidad de Washington. Intrigado por el tiempo desde su niñez, estudió su primer ciclo en ciencias físicas en la Universidad de California del Sur y después ingresó en la Fuerza Aérea como meteorólogo y observador en los vuelos de reconocimiento en el Pacífico Norte y Central. Tras dejar la Fuerza Aérea estuvo en el Instituto de Tecnología de Massachusetts, donde obtuvo su doctorado en meteorología en 1963. Durante los cinco años siguientes fue meteorólogo investigador en la Compañía Rand, que abandonó en 1968 para quedarse en la Universidad de Washington. Ha estado interesado en las atmósferas de los planetas desde hace más de diez años; en 1968 empezó a trabajar con Yale Mintz, de la Universidad de California en Los Angeles, sobre modelos de la dinámica de la atmósfera marciana. Desde entonces ha estado relacionado con la interpretación de los datos sobre los fenómenos de la atmósfera de Marte recogidos por los *Mariner 6, 7* y *9* y con los experimentos meteorológicos a bordo de los vehículos Viking que se posaron en Marte.

JOSEPH L. MELNICK, GORDON R. DREESMAN y F. BLAINE HOLLINGER ("Hepatitis vírica") pertenecen al claustro del Baylor College of Medicine. Melnick es profesor honorario de virología y epidemiología. Empezó investigando la hepatitis vírica durante la segunda guerra mundial, tras doctorarse en la Universidad de Yale, en cuya facultad de medicina permaneció 18 años, hasta ingresar en el Baylor. El año pasado dirigió el comité de la Organización Mundial de la Salud que investigaba el campo de la hepatitis vírica. Dreesman es profesor agregado de virología. Tras graduarse en el Central College de Iowa y licenciarse en la Universidad de Kansas se doctoró en microbiología por la Universidad de Hawaii en Manoa; centra su interés en la virología. Hollinger se doctoró en Kansas, especializándose en medicina interna. Tras dos años de trabajo, en el Center for Disease Control of the U.S. Public Health Service en Atlanta, y de investigación, como becario posdoctoral en virología fundamental bajo los auspicios de los National Institutes of Health, ingresó en el Baylor College, en 1970.

JOANDOMENEC ROS I ARAGONES ("La defensa en los opistobranquios") es profesor de ecología en la Universidad de Barcelona, en cuya ciudad nació. Doctor en ciencias biológicas (su tesis trata de la faunística y de la ecología de los opistobranquios ibéricos), buceador y apasionado por la biología marina, ha colaborado y participa en la actualidad en diversos programas de investigación sobre las comunidades bentónicas de las costas españolas, y se dedica especialmente al estudio de la ecología de los gasterópodos opistobranquios, sobre los que ha publicado varios trabajos.

D. D. DOUBLE y A. HELLAWELL ("La solidificación del cemento") son, el primero, becario de investigación en el departamento de metalurgia y ciencia de los materiales en la Universidad de Oxford y, el segundo, profesor de materiales en la Universidad de Wisconsin en Milwaukee. Double se graduó en Oxford, doctorándose con una tesis sobre la solidificación de los metales. A partir de entonces se dedicó al estudio de la hidratación del cemento portland. Hellawell también se doctoró en Oxford, pasando seguidamente a profesar metalurgia hasta el año pasado.

JAMES L. DYE ("Aniones de metales alcalinos") es profesor de química en la Michigan State University, donde trabaja desde 1953. Después de graduarse en el Gustavus Adolphus College, recibió su doctorado en química física de la Iowa State University en 1953. En 1961 ingresó como becario del National Science Foundation Science Faculty y llevó a cabo investigaciones en colaboración con Manfred Eigen en el Max Planck Institut de química biofísica en Göttingen. Ha estudiado las soluciones de metales alcalinos en amoníaco y en aminas desde 1954; trabaja subvencionado por la Energy Research and Development Administration (ERDA). Por otra parte, durante los últimos cinco años el doctor Dye ha empezado la investigación de los espectros ópticos de los compuestos intermedios formados durante las reacciones catalizadas por enzimas.

G. ADRIAN HORRIDGE ("El ojo compuesto de los insectos") profesa zoología en la Universidad de Cambridge, en cuyo centro universitario se formó, estudiando temas que iban desde las ramificaciones nerviosas de las medusas hasta los arrecifes coralinos del mar Rojo y las alfombras persas. Obtuvo su primer empleo en la Royal Aircraft Establishment en Farnborough, en donde trabajó en estructuras de ingeniería a base de materiales compuestos. Desde allí se trasladó a la Universidad de San Andrés, en donde posteriormente dirigiría el Gatty Marine Laboratory. En 1969 marchó a Australia, en cuya universidad central de Camberra fundó el departamento de neurobiología. Volvió a Cambridge el pasado febrero. Hoy, su entretenimiento principal consiste en el estudio de la construcción de los botes indonesios en los embarcaderos donde se construyen tradicionalmente tales veleros.

IAN STEWART ("Gauss") se encuentra en el instituto matemático de la Universidad de Warwick. Tras graduarse por la de Cambridge, se doctoró en ciencias exactas en la Universidad de Warwick. En 1974 estuvo en Tübingen becado por la fundación Humboldt. Su principal campo de investigación son las álgebras infinito-dimensionales de Lie. Ha publicado ocho libros sobre matemáticas (dos de ellos para niños) y está concluyendo su novena obra escrita.

Debate sobre el ADN recombinante

Se enmarcan, en una perspectiva más general, la controversia iniciada hace ya cuatro años sobre los peligros que podrían derivar de los métodos de obtención del ADN recombinante y la eficacia de las normas de contención establecidas

Clifford Grobstein

Las normas de investigación publicadas hace un año por los National Institutes of Health a propósito de la obtención de moléculas de ADN recombinante constituyeron la culminación de un esfuerzo extraordinario de autorregulación por parte de la comunidad científica. Sin embargo, la aparición de las normas de los NIH sobre la investigación en ADN recombinante no ha puesto punto final al debate. Antes al contrario, éste ha incrementado en los últimos meses en intensidad y en gama de procedencias de las personas participantes. En una sesión pública organizada en el pasado marzo por la Academia Nacional de Ciencias de Washington se puso de manifiesto el amplio abanico de opiniones. En dicha sesión se reafirmaron las posturas de los científicos que se habían establecido desde los comienzos del debate. Asistieron también nuevos participantes que presentaron un programa variado y dinámico. Dejaron bien sentado que la investigación del ADN recombinante se había convertido en un problema político. Tal como dijo uno de los conferenciantes, la discusión de la Academia había sido la discusión pública más importante sobre ADN recombinante organizada por los científicos implicados en esa línea de trabajo. Los no científicos asistentes al congreso insistieron, de palabra y obra, en que la ciencia ha adquirido dimensiones suficientes como para dejarla en manos de los propios científicos u otorgarle un velo de castidad o inmunidad política.

La ciencia implica, naturalmente, consecuencias decisivas para la sociedad, por la sencilla razón de que es fuente de beneficios y de riesgos. La investigación sobre ADN recombinante puede proporcionar nuevos e importantes beneficios sociales de magnitudes indefinidas: productos farmacéuticos más eficaces y baratos, mejor comprensión de las causas del cáncer, producción alimentaria más abundante e incluso nuevos enfoques respecto del problema de la energía. Se prevé estos y otros posibles resultados en los planes más optimistas sobre la aplicación futura de la tecnología del ADN recombinante. Pero también caben los "esquemas pesimistas": epidemias mundiales causadas por exponer a agentes patógenos de nueva creación, amenaza de catastróficos desequilibrios ecológicos, nuevas armas para militaristas y terroristas y fuerza para dominar y controlar el espíritu humano.

Los esquemas optimistas y los pesimistas se mueven en el terreno de la alta especulación: la distancia que los separa representa el elevado grado de incertidumbre que rodea a este gran salto adelante en la genética molecular. En las dos últimas décadas se ha fragmentado el material base de la herencia biológica, y parece como si cada uno de los fragmentos haya adquirido vida propia. En este período de inestabilidad resultante, el miedo amenaza con dominar el entusiasmo a medida que se van difundiendo las consecuencias implícitas en estas investigaciones. El miedo no lo es tanto de un peligro real y presente cuanto de futuros peligros imaginarios. La respuesta clásica a tales temores ha sido siempre un rígido muro de contención: la Gran Muralla, la Línea Maginot, la guerra fría. Todas ellas ponen de manifiesto el esfuerzo que se requiere para dar una seguridad absoluta ante riesgos impredecibles: cada una genera, a su vez, su propio riesgo. La intensificación del debate sobre el ADN recombinante tiene un componente de este tipo de comportamiento, pero existe otro componente nuevo más racional.

El primer episodio de este debate crucial data de 1974, cuando los investigadores pioneros en este campo acordaron aplazar voluntariamente varios tipos de experimentos que podían resultar arriesgados. Se había desarrollado un conjunto de técnicas que hacían posible fragmentar en segmentos más pequeños, con la ayuda de ciertos enzimas, los largos filamentos que constituyen las moléculas de ADN, los segmentos de ADN resultantes se recombinaban con el ADN de un vector apropiado, o transportador, y se reinsertaba el recombinante en la célula de un hués-

EL CODIGO GENETICO de un virus bacteriano extremadamente pequeño, el bacteriófago designado por ΦX174, viene dado por la secuencia de letras de la página opuesta. Las letras significan cuatro nucleótidos: citosina, guanina, adenina y timina; éstos se unen extremo a extremo para formar cada una de las dos cadenas de la molécula de ADN. El mensaje genético de cada cadena de ADN se encuentra en una secuencia determinada de nucleótidos, en la que un nucleótido puede venir precedido por cualquier otro. En el virus ΦX174 la molécula de ADN, que tiene solamente una única cadena circular durante parte de su ciclo biológico, está formada por 5375 nucleótidos, aproximadamente; los nucleótidos están agrupados en nueve genes conocidos, que son a su vez los responsables de codificar la secuencia de aminoácidos de nueve proteínas diferentes. Por ejemplo, el segmento de la molécula que tienen color oscuro, llamado gen *J*, codifica una pequeña proteína que forma parte del virus; resulta que este segmento también es el gen más corto del genoma del Φ X174. La secuencia de nucleótidos completa del ADN del Φ X174 fue obtenida recientemente por Frederick Sanger y sus colaboradores en el British Medical Research Council Laboratory of Molecular Biology, de Cambridge. Se necesitarían unas 2000 páginas de este tipo para poner de manifiesto la secuencia de nucleótidos del ADN del cromosoma de una bacteria unicelular típica; y se necesitarían aproximadamente un millón de páginas para exponer de forma similar el código genético de las moléculas de ADN que forman los cromosomas de una célula de mamífero.

GAGTTTTATCGCTTCCATGACGCAGAAGTTAACACTTTCGGATATTTCTGATGAGTCGAAAAATTATCTTGATAAACGAGGAATTACTACTGCTTGTTTA
TCAACTACCGCTTTCCAGCGTTTCATTCTCGAAGAGCTCGACGCGTTCCTATCCAGCTTAAAAGAGTAAAAGGCGGTCGTCAGGTGAAGCTAAATTAAGC
AACGATTCTGTCAAAAACTGACGCGTTGGATGAGGAGAAGTGGCTTAATATGCTTGGCACGTTCGTCAAGGACTGGTTTAGATATGAGTCACATTTTGTT
TGAACTGAGTACTAAAGAATGGATAATCACCAACTTGTCGTAGCCTGAGTCTATCATTAGGTGCGAGAAAATTTTACAGTTGTTCTCTTAGAGATGGTAC
TACTGAACAATCCGTACGTTTCCAGACCGCTTTGGCCTCTATTAAGCTCATTCAGGCTTCTGCCGTTTTGGATTTAACCGAAGATGATTTCGATTTTCTG
TCGCTCCCATAGGATGTTTCAGGTCGCATGGTATTTGCGTTCGGAGTTGCGTCGCTGCTCGTGCTCTCGCCAGTCATCGTTAGGTTTGAAACAATGAGCA
TTCCTGCTCCTGTTGAGTTTATTGCTGCCGTCATTGCTTATTATGTTCATCCCGTCAACATTCAAACGGCCTGTCTCATCATGGAAGGCGCTGAATTTAC
CAGTCATTCTTGCAGTCACAAAGGACGCGCATGTGCGTTCCATTTGCGCTTGTTAAGTCGCCGAAATTGGCCTGCGAGCTGCGGTAATTATTACAAAAGG
GCAGAAGAAAACGTGCGTCAAAAATTACGTGCGGAAGGAGTGATGTAATGTCTAAAGGTAAAAAACGTTCTGGCGCTCGCCCTGGTCGTCCGCAGCCGTT
TTAAATAGGAGTTCATTCCCCGGCTTCGGGGACGTTAATTTTAACAACTGGTGGATGTATGGTTTCTGCTCGCGGAAATGCGAACGGAAATCATGGAGCG
ATGTCTAATATTCAAACTGGCGCCGAGCGTATGCCGCATGACCTTTCCCATCTTGGCTTCCTTGCTGGTCAGATTGGTCGTCTTATTACCATTTCAACTA
TACAGATGTCATCTCAGTTATCGTTCCGGTGCTGCGTTACCTCTTTCTGCCTCTCGCGGTTGCCGCAGGTAGAGCTTCCTCAGCGGTCGCTATTGGCCTC
TTTTACTTTTTATGTCCCTCATCGTCACGTTTATGGTGAACAGTGGATTAAGTTCATGAAGGATGGTGTTAATGCCACTCCTCTCCCGACTGTTAACCAA
GATATCTATAGTTTATTGGGACTTTGTTTACGAATCCCTAAAATAACCATAGTCCCAATTAGCACGGTTCTTTTCGCCGTACCAGTTATATTGGTCATCA
CGTATTTTAAAGCGCCGTGG---ATGCCTGACCGTACCGAGGCTAACCCTAATGAGCTTAATCAAGATGATGCTCGTTATGGTTTCCGTTGCTGCCATCT
TCGTCGAACGTCTGGGTATTACAGTTATCTACACCATCTTCAGCAGTAAACCGCTCTTTCGAGTCAGAGTCCTCCTTCGCCTCGTCAGGTTTACAAAAAC
TATGCTAATTTGCATACTGACCAAGAACGTGATTACTTCATGCAGCGTTACCATGA-GTTATTTCTTCATTTGGAGGTAAAACCTCATATGACGCTGACA
ACTTGTGCTGGTCTTTTGACCGGATTGCTGCAAACCAGTCAAGGTAGTTGTAGTATCGGTCTACGGGTCTCTAATCTCGCGTACTGTTCATTTCCTGCCA
ACAGACCTATAAACATTCTGTGCCGCGTTTCTTTGTTCCTGAGCATGGCACTATGTTTACTCTTGCGCTGGTTCGTTTTCCGCCTACTGCGACTAAAGAG
GGAAGTATCTTTAAAGTGCGCCGCCGTTCAACGGTATGTTTTGTCCCAGCGGTCGTTATAGCCATATTCAGTTTCGTGGAAATCGCAATTCCATGACTTA
ATGTTTTCCGTTCTGGTGATTCGTCTAAGAAGTTTAAGATTGCTGAGGGTCAGTGGTATCGTTATGCGCCTTCGTATGTTTCTCCTGCTTATCACCTTCT
TTGACTTGCTGACTTTGTGACCAGTATTAGTACCAACGCTTATTCATGCGCAAGAACGTTTAGTGGTCTTCCGCCAAGGACTTACTTACCCTTCGGAAGT
GTTGCAGTGGATAGTCTTACCTCATGTGACGTTTATCGCAATCTGCCGACCACTCGCGATTCAATCATGACTTCGTGATAAAAGATTGAGTGTGAGGTTA
TTTCAGACTTTGTACTAATTTGAGGATTCGTCTTTTGGATGGCGCGAAGCGAACCAGTTGGGGAGTCGCCGTTTTTAATTTTAAAAATGGCGAAGCCAAT
TATTTCTCGCCACAATTCAAACTTTTTTTCTGATAAGCTGGTTCTCACTTCTGTTACTCCAGCTTCTTCGGCACCTGTTTTACAGACACCTAAAGCTACA
GGACTAATCGCCGCAACTGTCTACATAGGTAGACTTACGTTACTTCTTTTGGTGGTAATGGTCGTAATTGGCAGTTTGATAGTTTATATTGCAACTGCT
TTGTTTCAGTTGGTGCTGATATTGCTTTTGATGCCGACCCTAAATTTTTTGCCTGTTTGGTTCGCTTTGAGTCTTCTTCGGTTCCGACTACCCTCCCGAC
TAACGGCCCGCATGCCCCTTCCTGCAGTTATCAGTGTGTCAGGAACTGCCATATTATTGGTGGTAGTACCGCTGGTAGGTTTCCTATTTGTAGTATCCGT
AACGTCTACGTTGGTTTCATGGTTTGGTCTAACTTTACCGCTACTAAATGCCGCGGATTGGTTTCGCTGAATCAGGTTATTAAAGAGATTATTTGTCTCC
AACTGGCGGAGGTTTGTTAAATCTGTACCGCGGTGGTCGTTCTCGTCTTCGTTATGGCGGTCGTTATCGTGGTTTGTATTTAGTGGAGTGAATTCACCGA
AAAGCCGCCTCCGGTGGCATTCAAGGTGATGTGCTTGCTACCGATAACAATACTGTAGGCATGGGTGATGCTGGTATTAAATCTGCCATTCAAGGCTCTA
CTTCACGGTCGGACGT GCATGGAAGTTCTTCAGGAAATGGTCGAAATCGTTATCGTGTGCTTTGTTTTGATCCCCGCCGGAGTAGTCCCAATCCTTGTA
TGCCGTTTCTGATAAGTTGCTTGATTTGGTTGGACTTGGTGGCAAGTCTGCCGCTGATAAAGGAAAGGATACTCGTGATTATCTTGCTGCTGCATTTCCT
GGTCAACGTAAAATCATTCGAGAAAAACTAAGAGTTTAGGCCGCAGTTGGTATGGTCGTCTCCTTCGTAGTCGTGGTCGTGCGAGGGTTCGTAATTCGAG
ACAATCAGAAAGAGATTGCCGAGATGCAAAATGAGACTCAAAAAGAGATTGCTGGCATTCAGTCGGCGACTTCACGCCAGAATACGAAAGACCAGGTATA
CCTTTGGACGACAACGAACCTTTCTAACCACAAAAGGTATTATCTGCGTTGCGTCGTCATCTGAGGAAGACA-CTTATTCGTTCGTAGAGTAAAACACGT
GAGATTATGCGCCAAATGCTTACTCAAGCTCAAACGGCTGGTCAGTATTTTACCAATGACCAAATCAAAGAAATGACTCGCAAGGTTAGTGCTGAGGTTG
GGTCTTCGTCGTAGTCACTGCTGTAATCTTTATAGGAAACGTCATCGCGGTTATACTCTTCTCGGTATGGCGACTAAGACGCAAACGACTACTTGATTCA
TGTGGTTGATATTTTTCATGGTATTGATAAAGCTGTTGCCGATACTTGGAACAATTTCTGGAAAGACGGTAAAGCTGATGGTATTGGCTCTAATTTGTCT
TAAGTCTTCCCATTATTCTTGCTTGGTATTTTTTCGGAGGTTCTAAACCTCCGTACTTTTGTATGTTAACCCTCCCACAGTTAGGACTGCCAATAAAGGA
GTCACGCTGATTATTTTGACTTTGAGCGTATCGAGGCTCTTAAACCTGCTATTGAGGCTTGTGGCATTTCTACTCTTTCTCAATCCCCAATGCTTGGCTT
CGGCAGTTGTATGTATAGTGGTAATAGCTTGAGTTGCGGGACGTATGCTTTTCTGTCTTAGAGAAGGTTCTCGAACTACGCCAATAGGTAGACGAATACC
CATAAGGCTGCTTCTGACGTTCGTGATGAGTTTGTATCTGTTACTGAGAAGTTAATGGATGAATTGGCACAATGCTACAATGTGCTCCCCCAACTTGATA
CCCGCAAGTCGTCGGTCGAACGTTTTGACGCATTGGCAGAAGAGCAAGAGATTTTTGGTAAAAAGCAGGGGAAGCCCCGCCACCAGATATCACAATAATT
TCTTAAGGATATTCGCGATGAGTATAATTACCCCAAAAAGAAAGGTATTAAGGATGAGTGTTCAAGATTGCTGGAGGCCTCCACTAAGATATCGCGTAGA
GATTAGCCAGCAGTCGGTTGCACTCTCACAGTTTTTGCTATTTGGTTGGTAGTCGTACTCGGACAGCGTAACGTAAGTAGTTTGCGACTTATCGTTTCGG
AGGCGTTTTATGATAATCCCAATGCTTTGCGTGACTATTTTCGTGATATTGGTCGTATGGTTCTTGCTGCCGAGGGTCGCAAGGCTAATGATTCACACGC
CGATGGACATCCTTCACAGGCGTATTTCACGTGGCGTACCTTTACTTCTGCCGGTAATCGACATGGTATGAGTCCGTGTGTTTTTATGACTATCGTCAGC
GTTGACCCTAATTTTGGTCGTCGGGTACGCAATCGCCGCCAGTTAAATAGCTTGCAAAATACGTGGCCTTATGGTTACAGTATGCCCATCGCAGTTCGCT
TGTATCTTTGGTTGTCGGTATATTGACCATCGAAATTCGCCGAGTGGAAATCGTAGTTGTCCGGTGTTGGTTGGTCTTGCACTTTTTCGCAGGACGCACA
GGCTAAATACGTTAACAAAAAGTCAGATATGGACCTTGCTGCTAAAGGTCTAGGAGCTAAAGAATGGAACAACTCACTAAAAACCAAGCTGTCGCTACTT
CGAACCATTCAACCTAATTCGTGAGGCACCTGTCTAAACAGTAACACTCGTAAAAGTAGGGCTTCAACGCCGAGTAAGACTAAGACTTGTCGAAGAACCC
TGGGTTACGACGCGACGCCGTTCAACCAGATATTGAAGCAGAACGCAAAAAGAGAGATGAGATTGAGGCTGGGAAAAGTTACTGTAGCCGACGTTTTGGC
ACGTCCAACCTATGCGGTTAGTAAAAATAGCTTCGCGCGTATTTAAACTCGTCTAAACAGCAGTGTCCAACGCGG

ped adecuado con el fin de que se propagara, y, de ser posible, se mostrara capaz de funcionar.

El significado de los nuevos descubrimientos radica en el papel biológico central del ADN como transmisor de la información genética de generación en generación. La transmisión del mensaje genético codificado depende de la capacidad de una célula para generar réplicas exactas del ADN parental y distribuir dichas réplicas entre las células hijas. El éxito de la transmisión genética depende, además, de la capacidad de las células hijas para "expresar" correctamente la información codificada, recurriendo a ella para controlar los procesos vitales esenciales.

El mecanismo de la expresión genética en los organismos superiores sólo se conoce vagamente, y se ha supuesto que el descubrimiento de las nuevas técnicas de recombinación del ADN ha de comportar la profundización inmediata en los conocimientos en este campo.

Los mecanismos de la replicación y expresión genética son enormemente complejos. Sin embargo, la esencia de la cuestión se halla en la famosa estructura en "doble hélice" del ADN. Las dos cadenas de la molécula de ADN, largas, arrolladas y complementarias, están formadas por cuatro tipos de nucleótidos, a saber: citosina, guanina, adenina y timina (abreviadamente, *C, G, A y T*) que se engarzan extremo a extremo como los vagones de un tren. El mensaje genético de cada cadena se encuentra en una secuencia determinada de nucleótidos en la que cada uno puede venir precedido por cualquier otro de ellos. Por ejemplo, la secuencia *CATTACTAG* contiene cinco palabras inglesas identificables: *CAT, AT, TACT, ACT y TAG*. Sin embargo, el mensaje genético "está escrito" en tripletes: *CAT, TAC y TAG*. En general, cada triplete "codón" determina, a través de una serie de pasos intermedios, la posición de un aminoácido particular en una molécula proteica.

Al igual que los ácidos nucleicos, las proteínas pueden representarse como largos trenes formados por vagones unidos extremo a extremo: sin embargo, en este caso las subunidades son aminoácidos en lugar de nucleótidos. La secuencia de nucleótidos de una molécula dada de ADN determina la secuencia de aminoácidos de una proteína particular, colocando cada codón, tri-nucleótido, a cada uno de los 20 aminoácidos posibles en posiciones sucesivas de la cadena proteica. A su vez, la secuencia de aminoácidos determina de modo específico la estructura y la función de la proteína. Por tanto la secuencia de nucleótidos del ADN determina la capacidad de síntesis de proteínas por parte de un organismo. Además virtualmente cada propiedad del organismo, desde la acción enzimática al color de los ojos, depende de un modo u otro de la estructura proteica.

La transmisión de generación en generación de la información genética esencial depende de la replicación exacta de la secuencia de nucleótidos del ADN. El mecanismo de replicación radica en la relación complementaria entre las dos cadenas de la molécula de ADN. Una secuencia de una cadena (por ejemplo *CATTACTAG*) se encuentra opuesta a una secuencia complementaria (*GTAA-TGATC*) de la otra cadena. Las cadenas son complementarias porque *C* y *G* se encuentran siempre en posiciones opuestas en las dos cadenas de ADN, al igual que *A* y *T*. La complementariedad depende de la afinidad química particular, o unión, entre *C* y *G* por una parte y *A* y *T* por otra. La suma de estos enlaces, repetidos a lo largo de las cadenas es lo que las mantiene unidas, formando la doble hélice. En condiciones apropiadas se reduce la afinidad y las dos cadenas pueden desenrollarse y separarse. Las cadenas sencillas pueden aparearse y enrollarse de nuevo cuando se restablecen las condiciones en que se produce una afinidad elevada.

El ADN de doble cadena se replica gracias a una extensión de estas propiedades. El desenrollamiento y separación de las dos cadenas empieza en un lugar localizado a lo largo de la molécula de ADN. En presencia de enzimas apropiados y de nucleótidos libres, se forma una nueva cadena junto a la parte expuesta de cada cadena vieja desapareada. Cada nucleótido se alinea junto a su opuesto (*C* junto a *G*, *A* junto a *T*). La secuencia complementaria establecida así se une después, extremo a extremo, merced al concurso de un enzima que enlaza los nucleótidos individuales. Cuando el proceso de replicación se ha extendido, finalmente, por toda la doble hélice original, se han formado ya dos nuevas hélices idénticas a la primera. La replicación del ADN es la reacción química más importante del mundo vivo. Corrobora, cabalmente, el primer principio clásico de la herencia: una copia genera otra copia idéntica.

Si la replicación del ADN se realizara siempre sin error, la vida sería mucho más homogénea de lo que es. Y aquí es donde interviene el segundo principio clásico de la herencia: el principio de la variación mutacional o la aparición en la progenie de nuevas características hereditarias que no estaban presentes en los progenitores. Las mutaciones aparecen, al menos parcialmente, por errores en el proceso de replicación. Por ejemplo, la sustitución de un nucleótido por otro cambia el triplete que forma el codón y coloca un aminoácido diferente en la posición correspondiente en la proteína resultante. Los errores en un nucleótido repercuten en un aminoácido. Así, el error en un solo nucleótido es el responsable de la anemia falciforme. La mayoría de las mutaciones no son tan simples como el mero intercambio de un solo nucleótido; no obstante, están relacionadas directamente con alteraciones, transposiciones o deleciones de la secuencia de nucleótidos del ADN.

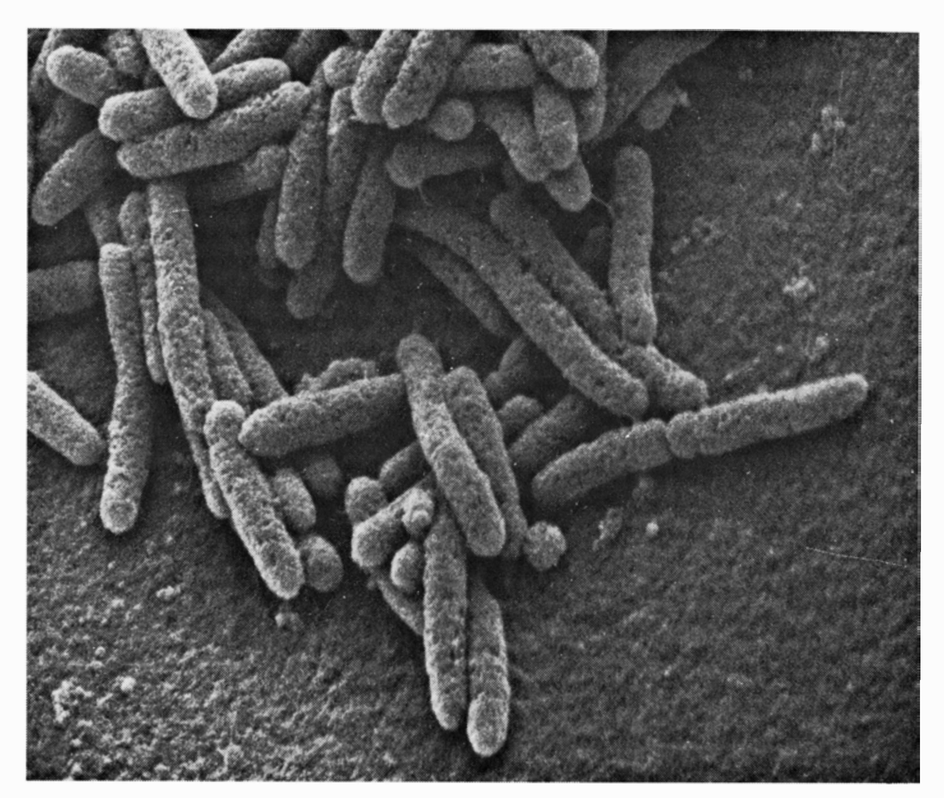


BACTERIOFAGO ΦX174 CON SU ADN. El virus infecta a la bacteria intestinal *Escherichia coli*. En la microfotografía de la izquierda, realizada por Jack D. Griffith, de la Facultad de Medicina de la Universidad de Stanford, y Andrew Stechelin, de la Universidad de Colorado, se pueden observar dos partículas ΦX174 unidas a la superficie de una célula de *E. coli*. En la microfotografía de la derecha, realizada por Griffith, pueden observarse las moléculas de ADN de dos virus ΦX174 en su forma de doble cadena; cada molécula tiene una longitud aproximada de 18.000 angstroms.

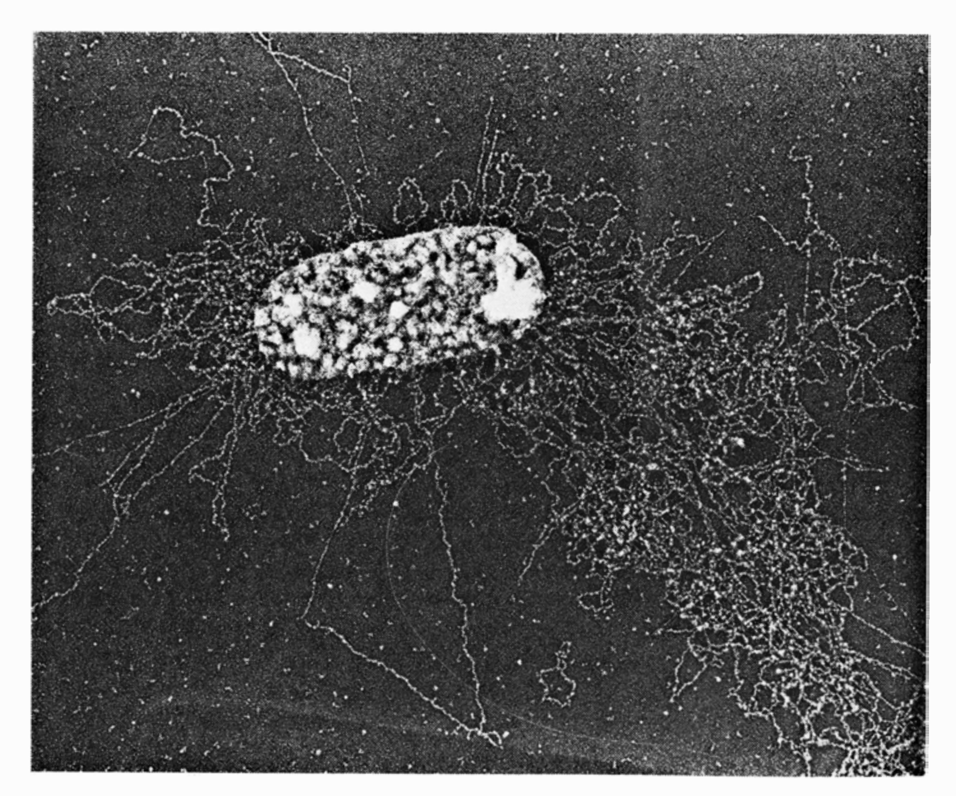
Cuando estos cambios aparecen en un gen (que es un segmento de ADN que codifica la secuencia de aminoácidos de una proteína determinada), producen cambios correlativos en la proteína y, por tanto, en las propiedades hereditarias que controla.

En esto radica la importancia de la tecnología del ADN recombinante. Por primera vez, hace posible la manipulación directa de las secuencias de nucleótidos. Los cambios en las secuencias de nucleótidos causados por "errores" naturales se producen al azar, incluso cuando su frecuencia total se incrementa artificialmente. En las poblaciones naturales, la selección darwiniana "escoge" entre los errores producidos al azar, incrementando la representación de aquellos que dan lugar a una mayor descendencia en un ambiente determinado. La selección artificial, que ha sido practicada por los seres humanos desde hace milenios, favorece los errores que satisfacen las necesidades humanas (producción agrícola) o sus caprichos (cría de animales exóticos). Sin embargo, tanto la selección natural como la artificial dependen de que se produzcan al azar las mutaciones deseadas. Hasta que surgieron las técnicas de obtención del ADN recombinante no existía ninguna forma de dirigir el cambio genético. Estas nuevas técnicas permiten introducir deliberadamente secuencias de nucleótidos conocidas y adecuadas de una cepa o una especie en otra, confiriéndole de este modo una propiedad deseada.

El estudio del ADN recombinante exige ingenio experimental y conocimiento detallado de la molécula de ADN. Empieza por un ataque del ADN por las proteínas denominadas enzimas de restricción, que se aíslan a partir de las bacterias. El ataque enzimático rompe la doble cadena de ADN reconociendo secuencias determinadas, por ejemplo, la secuencia *CATTAC* cuya complementaria es *GTAATG.* La rotura no ocurre siempre en el mismo punto en las dos cadenas. Por ejemplo, puede tener lugar entre las dos *T,* de la primera cadena, y justo a la derecha de *ATG* en la segunda. Por tanto, al separarse, un fragmento termina en *TAC,* mientras que el otro lo hace en *ATG.* Ya que los extremos de la cadena sencilla son complementarios, en condiciones apropiadas permanecerán unidos uno al otro y podrán enlazarse sus extremos. Si se utiliza el mismo enzima de restricción sobre un ADN que proceda de dos fuentes diferentes, que tengan ambas la secuencia objeto apropiada, entonces resultarán secuen-



CONJUNTO DE CELULAS DE *E. COLI.* Microfotografía electrónica realizada por David Scharf. Unas células se hallan en el momento de su reproducción asexual (división celular); otras, según parece, están transfiriendo su ADN a través de la conexión filamentosa característica del proceso conocido como conjugación. La mayoría de los investigadores consideran que la bacteria *E. coli* es la célula huésped más apropiada para los experimentos de recombinación del ADN.



***E. COLI* DEJA SALIR EL ADN a través de su membrana celular fragmentada, en esta microfotografía electrónica, realizada por Griffith. La mayor parte del ADN está formado de una única y larga molécula de ADN de doble cadena, que constituye el cromosoma de este organismo procariota simple. Además, la bacteria *E. coli* puede tener uno o más anillos de ADN, que se replican inmediatamente, conocidos como plásmidos; en la parte baja de la microfotografía puede verse una de estas moléculas de ADN extracromosómico, más pequeñas. Los plásmidos, procedentes de las células de *E. coli*, desempeñan un papel importante en la investigación sobre el ADN recombinante, ya que constituyen una clase de vectores, o transportadores, a los que se pueden unir segmentos de ADN "extraño" antes de reinsertarse en una célula huésped adecuada para propagarse, duplicando la propia secuencia original de nucleótidos y la secuencia extraña.**

cias con idénticos extremos "pegajosos". Aprovechando esta unión, dos secuencias procedentes de cualquier fuente pueden recombinarse en una única molécula de ADN.

La etapa siguiente, y última, a recorrer será colocar el ADN recombinante en un organismo huésped apropiado. El recombinante debe poder penetrar en el huésped y llegar a formar parte de su sistema genético. Se ha desarrollado un método eficaz para conseguir esto con la enterobacteria *Escherichia coli.* Además de su único cromosoma, largo y circular, la bacteria *E. coli* puede tener uno o más anillos pequeños de ADN que se replican independientemente; éstos reciben el nombre de plásmidos. Los plásmidos pueden ser aislados a partir de esas bacterias, fragmentados por acción de los enzimas de restricción y ser utilizados como un componente de un recombinante. Después de enlazar ADN del plásmido con ADN "extraño", el plásmido recupera su forma circular y se obtiene de nuevo una célula completa. Allí se inicia nuevamente la replicación, duplicándose no sólo la propia secuencia original sino también la extraña. Puede obtenerse así una cepa bacteriana que sintetizará un número indefinido de copias de la secuencia de nucleótidos insertada, procedente de otro organismo.

Hasta aquí, nada parece ser particularmente grave o alarmante; se trata únicamente de una química nueva y apasionante aplicada a los organismos vivos. Sin embargo, dada la complejidad de éstos y del mundo aún más complejo de los fenómenos sociales, esta nueva química descubre posibilidades muy diversas, teóricas y reales. Supongamos, por ejemplo, que se aislara la secuencia de nucleótidos necesaria para sintetizar una toxina potente y se insertase en *E. coli,* que es un organismo, inocuo de ordinario, que se encuentra en el intestino humano. ¿Surgiría un nuevo y peligroso agente patógeno? ¿Liberaría la *E. coli* transformada alguna toxina en el intestino humano? ¿Podría dicho nuevo agente patógeno escapar de los mecanismos de control y producir epidemias? Todas esas cuestiones tienen su respuesta, pero ha de transcurrir cierto tiempo antes de que las conozcamos. A fin de disponer de tiempo para reflexionar, los investigadores pidieron en 1974 un aplazamiento parcial y temporal de los experimentos que se consideraron potencialmente más arriesgados. El propósito primordial de las normas publicadas por los NIH, en junio de 1976, fue deslindar los experimentos seguros, exentos de peligro, de los más arriesgados. Las normas, que substituyeron al aplazamiento temporal, tomaban como punto de arranque los análisis más pesimistas de distintos tipos de experimentos; tuvieron por objeto valorar el alcance presumible de los peligros y emitir normas de defensa equivalentes y adecuadas para reducir los riesgos desconocidos.

Si se ha adoptado un código de reglas ¿por qué continúa el debate? Simplemente porque al comparar los riesgos con las medidas de protección dictadas como las normas de trabajo, los críticos las consideran inadecuadas, si se tienen en cuenta los posibles daños biológicos, e incompletas, al olvidar otros puntos importantes. Los críticos más elocuentes han presentado sus propios análisis pesimistas en la prensa especializada y general. Estos informes han extendido la alarma y han provocado, en los Estados Unidos, comentarios públicos a nivel local, estatal y federal. La preocupación que se ha ido expresando ha creado la necesidad de que se establezca un nuevo calendario para revisar la política a seguir en una segunda ronda de negociaciones.

Continúa dominando el debate el tema de los posibles daños biológicos y el grado de riesgo calculado. Las normas de los NIH equilibran el riesgo esperado en un experimento determinado y recomiendan medidas específicas para contener tales riesgos. (Riesgo, debe recordarse, significa posible peligro, no peligro demostrado.) Los experimentos que se supone presentan un riesgo excesivo quedan absolutamente proscritos. En el extremo polar, a los experimentos que se considera que presentan un riesgo insignificante se les exige solamente que cumplan las medidas precautorias habituales en un trabajo cuidadoso de laboratorio. Entre tales extremos, las normas establecen varios niveles de riesgo y prescriben, mediante combinaciones adecuadas, un incremento gradual en punto a protección física y biológica. Queda prohibido liberar al ambiente cualquier organismo recombinante.

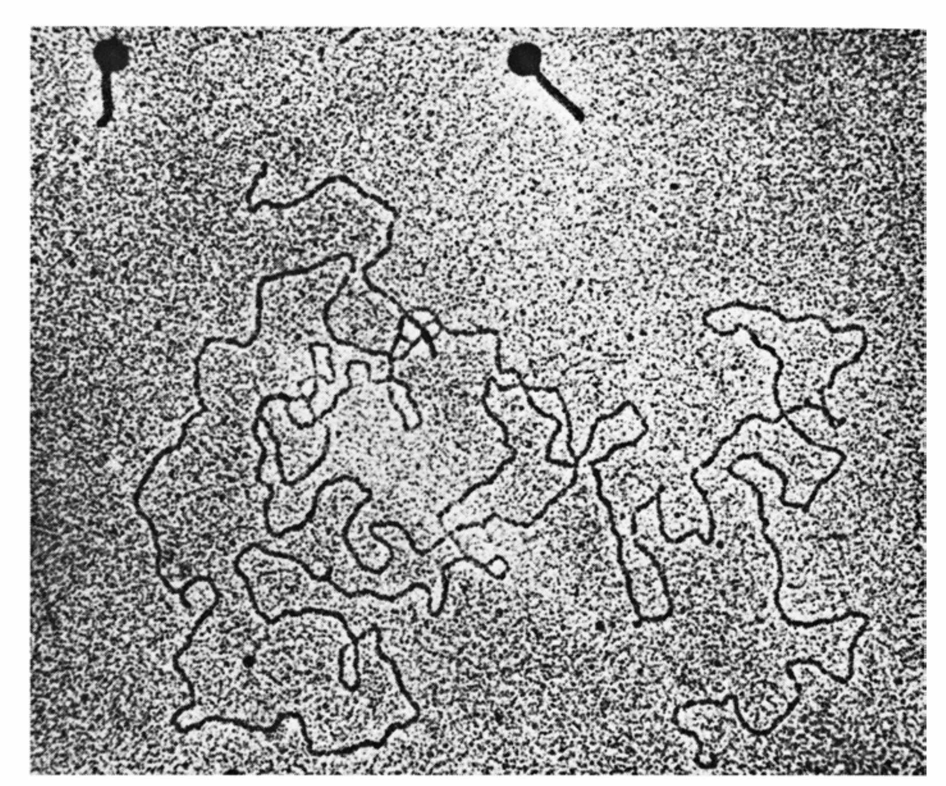
Desgraciadamente visto el limitado estado de conocimientos, a pesar de su progreso creciente, cabe el que se dé un profundo desacuerdo tanto en lo que se refiere a los grados de riesgo estimados como por lo que respecta a la eficacia de las medidas de contención propuestas. Unos, críticos, dirigen su información fragmentaria hacia la inevitable difusión de los nuevos y peligrosos organismos creados, que amenazan a la salud pública y al ambiente. Otros, defensores, dirigen la misma información fragmentaria hacia la conclusión de que las normas de los NIH pecan de cautela excesiva; creen que los posibles peligros actuales, con las medidas de precaución que se están tomando, no son mayores que los que afrontamos diariamente con los automóviles, aviones y otros servicios técnicos aceptados. Habida cuenta de la multiplicidad de los experimentos imaginables y de que la experiencia y datos críticos sobre

EL VIRUS BACTERIANO ESTA UNIDO a la pared de una célula de *E. coli*, en esta microfotografía electrónica realizada por Maria Schnoss, de la Facultad de Medicina de Stanford. Este virus, denominado bacteriófago lambda, suele infectar a la bacteria inyectando su ADN en la célula huésped por medio de un largo apéndice en forma de cola. El aumento es de unos 140.000 diámetros.

muchos puntos resultan insuficientes, no extraña ese abanico de opiniones. Un hecho es cierto: todavía no se ha producido ningún acontecimiento funesto en la investigación del ADN recombinante.

De todo lo cual se deduce una conclusión, a saber: la necesidad de llevar una investigación orientada eficazmente a reducir el grado de duda actual acerca del riesgo de determinadas clases de experimentos. Existe, por ejemplo, desacuerdo sobre el uso de *E. coli* como huésped del ADN recombinante. Unos sostienen que los científicos deben haberse vuelto locos para escoger un organismo que habita normalmente en el hombre a veces con actividad patógena como huésped del ADN recombinante. Este punto de vista, llevado al extremo, postula la suspensión de toda investigación en torno al ADN recombinante, mientras no se encuentre otro organismo más seguro. Otros autores argumentan: (1) que la gran cantidad de información existente sobre *E. coli* le confiere un valor incalculable; (2) que la cepa *K*-12 de *E. coli* usada actualmente en los laboratorios de investigación ha sufrido tal modificación genética, para adaptarla a las condiciones de laboratorio, que sólo con dificultad sobrevive en el intestino humano, y (3) que se han desarrollado nuevas cepas de *K*-12 con deficiencias genéticas adicionales que hacen inviable su supervivencia fuera de las condiciones del laboratorio. El uso de tales cepas genéticamente deficientes es lo que se entiende por "contención biológica". Apoyan esa idea quienes proponen este tipo de investigación como enfoque nuevo y eficaz que garantiza una mayor seguridad, y la rechazan quienes sostienen que puede quedar burlada por la recombinación natural.

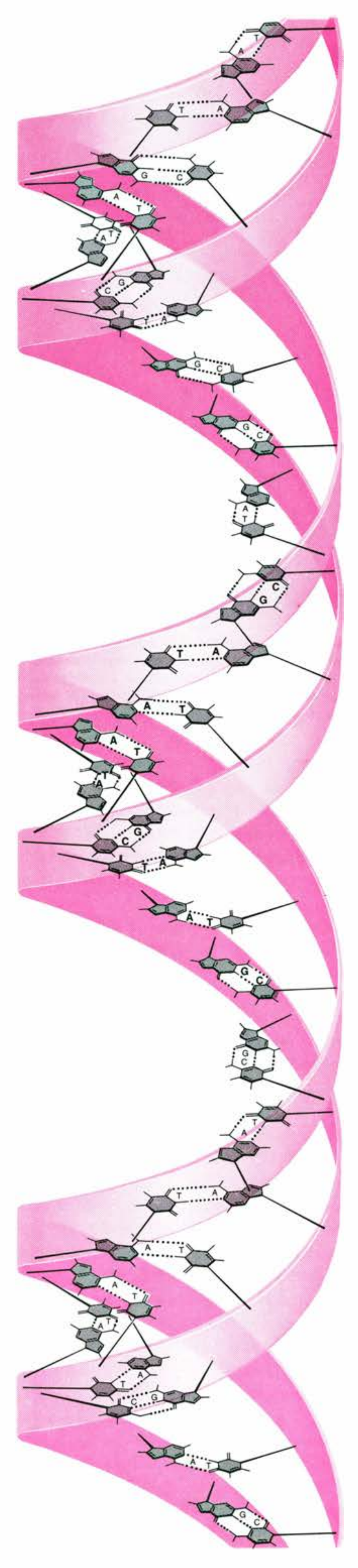
Los científicos suelen reducir esas divergencias de opinión a prácticas experimentales. Por ejemplo, la oportunidad de usar, como organismo experimental, la cepa *K*-12 de *E. coli* sólo puede juzgarse a partir del efecto de los genes recombinantes en las relaciones ecológicas de *E. coli* en el interior del intestino humano, cuyo efecto comprende el éxito de las cepas recombinantes en su competición con otras cepas de *E. coli* y otros organismos. Va disponiéndose de mayor información al respecto. Sin embargo, tales cuestiones no suelen ser temas de profundo interés científico. Se han convertido recientemente en materias prioritarias por la única razón de que pueden aportar una base de conocimientos útil para decidir los criterios a seguir. La planificación de programas o líneas de trabajo no ha sido nunca muy apreciada por parte de los científicos dedicados a la investigación básica, ni siquiera por los que se dedican a la investigación aplicada. El estudio de los criterios que deben regir los programas de investigación, consiguientemente, debe ser promovido mediante ayudas económicas especiales y a través de nuevos y adecuados canales institucionales. Se necesita urgentemente un centro, no dependiente de los NIH, que regule la investigación sobre el ADN recombinante y sobre otros trabajos que encierren un peligro potencial, y que podría disponer además de una sección dedicada a la investigación. El Center for Desease Control y su National Institute for Occupational Safety and Health, de los Estados Unidos, son posibles modelos de un centro destinado a esa doble misión.



EL BACTERIOFAGO LAMBDA Y SU ADN están ambos representados en esta microfotografia electrónica cedida por Griffith. En la parte superior hay dos virus lambda completos; debajo de ellos, está la molécula de ADN larga, de doble cadena, de un lambda fragmentado. La molécula de ADN del bacteriofago lambda puede utilizarse también como vector en los experimentos de recombinación del ADN en que intervienen células que desempeñan misión hospedante. *E. coli.*

En relación con la cuestión de los posibles peligros se discute también la oportunidad de centralizar las ayudas para la investigación sobre ADN recombinante. Algunos de los que temen que los organismos recombinantes produzcan graves daños han sugerido que la investigación potencialmente más peligrosa se concentre en lugares aislados y se apliquen procedimientos de contención extremadamente rigurosos. Los que minimizan el peligro se oponen a la idea de concentración, porque, sostienen, tendería a separar la investigación de la corriente intelectual y resultaría innecesariamente cara en cuanto a la obtención de medios de trabajo. Este argumento ha sido particularmente enérgico respecto a los experimentos que precisan medios de trabajo del tipo *P*3, que se definen como los necesarios para contener los experimentos con "riesgos moderados". Se supone que serán menos los equipos de investigación que requieran medios del grado *P*4, clave para designar la experimentación que implica "riesgos elevados", dado su alto coste implicado; además, es probable que resulten relativamente aislados. Las normas actuales de los NIH nada o poco orientan sobre estos extremos. En la orden del día del nuevo calendario se ha incluido el estudio de la posible aplicación de una política más dura y exigente. En particular es urgente la atenta consideración de ciertas posibilidades intermedias, como el uso de medios de trabajo de riesgo elevado, centralizados, para obtener por primera vez determinadas recombinaciones. Estas actividades, junto con las pruebas preliminares de los posibles peligros de los nuevos recombinantes, también deberían discutirse en el nuevo calendario.

Un caso especial que refuerza la importancia de realizar las pruebas iniciales



en centros especialmente preparados es el de los experimentos llamados de "perdigonada". Estos experimentos, que ofrecen especiales ventajas para el investigador, pueden presentar también peligros especiales. En los experimentos de perdigonada se expone todo el ADN de un organismo a los enzimas de restricción para obtener muchos fragmentos de ADN. Cada uno de estos fragmentos se recombina luego con ADN de un vector apropiado, y los recombinantes se vuelven a enlazar al azar al ADN de células huéspedes de *E. coli.* El paso siguiente es extender las células de *E. coli* ante un sustrato nutritivo, a fin de que cada célula receptora, que contiene una secuencia extraña determinada forme una colonia. Si el experimento tiene éxito, el resultado es una "individualización" de todas las secuencias de nucleótidos de un organismo determinado, creciendo cada una de ellas en una cepa separada y accesible a la manipulación y a la recombinación cruzada.

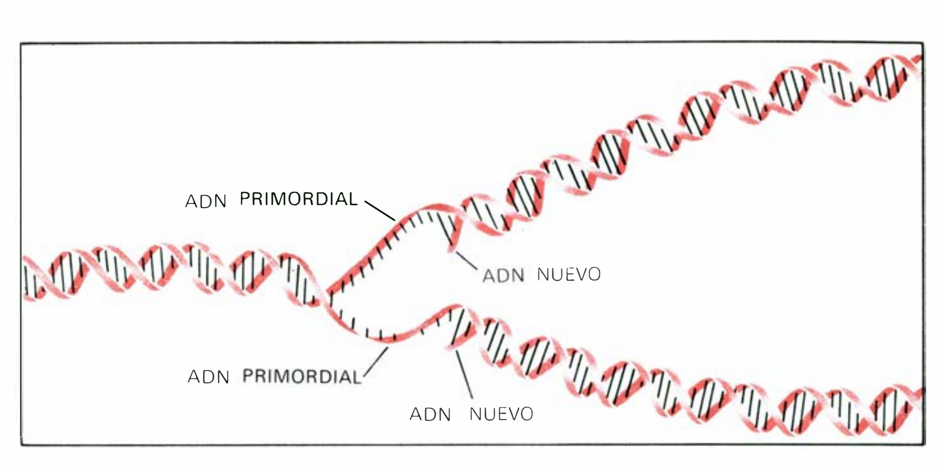
Este ensayo es laborioso, pero menos que cualquier otro que pueda ser útil para el estudio de los sistemas genéticos complejos de los organismos superiores. Existe, sin embargo, un riesgo, de magnitud desconocida, de que los fragmentos de ADN de función ignorada o no manifestada puedan duplicarse y causar peligros no previstos. El resultado podría ser catastrófico si la preparación de ADN original contuviera material genético de parásitos o virus asociados a las especies en estudio. Por tanto, de acuerdo con las normas de los NIH, los experimentos de perdigonada se consideran más peligrosos que aquellos otros en los que interviene ADN purificado y caracterizado. Los experimentos de esta categoría se consideran tanto más peligrosos cuanto más próximo sea biológicamente el organismo en estudio al organismo humano. Así se da por sentado que los experimentos con ADN de primate resultan más peligrosos que los experimentos con ADN de ratón. Esta consideración parece ser una medida de precaución razonable en lo que se refiere a los posibles peligros para la salud humana, pero no lo es tanto respecto a los posibles efectos ecológicos. Por ejemplo, los recombinantes obtenidos con la técnica de perdigonada en los que interviene ADN procedente de plantas podrían producir daños ecológicos si escaparan al ambiente. Por consiguiente, es mejor que los experimentos de perdigonada se realicen primero en centros especiales, con medios adecuados, que, además, podrían servir como lugares de almacenamiento y distribución de los productos recombinantes una vez comprobada su inocuidad.

Estos ejemplos ponen de manifiesto las ventajas que reportaría la creación de un Centro de Recursos Genéticos. En ese centro no sólo se deberían llevar a cabo las recombinaciones de ADN que pudieran ser peligrosas, sino que podría encargarse además de preservar la información genética contenida en especies naturales que se ven amenazadas y la contenida en cepas especiales de células u organismos obtenidos artificialmente para la investigación y otros fines. Es lógico suponer que la información genética almacenada adquirirá importancia en el futuro. En este sentido, se usan desde hace tiempo mezclas genéticas en cepas domésticas de plantas y animales, procedentes de progenitores salvajes, para fortalecer la respuesta de las cepas domésticas frente a las condiciones cambiantes de su cultivo y cuidado. Los progenitores salvajes se ven amenazados por la reducción a escala mundial de los hábitat salvajes.

La posibilidad de un daño biológico no surge necesariamente como consecuencia de la investigación básica. Las aplicaciones prácticas de las técnicas del ADN recombinante junto con la investigación aplicada y el desarrollo que ha conducido a ellas son fuentes por lo menos igualmente probables. Por ejemplo, las técnicas de recombinación pueden difundir enormemente el uso de las bacterias (y otros microorganismos) para la producción de ciertas proteínas y

LA ESTRUCTURA EN DOBLE HELICE DEL ADN queda de manifiesto en este diagrama simplificado de un segmento corto de la molécula de ácido desoxirribonucleico (ADN). El azúcar y los grupos fosfato que se unen extremo a extremo para formar la estructura externa de la molécula de doble cadena, se representan aquí esquemáticamente por las dos bandas helicoidales coloreadas. La porción interna de cada cadena de polinucleótidos, dibujada con mayor detalle, está formada por una secuencia variable de cuatro tipos de bases: dos purinas (adenina y guanina, *A* y *G*) y dos pirimidinas (timina y citosina, o *T* y *C*). Las dos cadenas, de direcciones opuestas, se mantienen unidas por enlaces de puentes de hidrógeno (*líneas negras partidas*) entre pares de bases. La adenina se aparea siempre con la timina y la guanina se aparea siempre con la citosina. Los planos de las bases son perpendiculares al eje común a las dos hélices. El diámetro de la doble hélice mide 20 angstroms. Las bases adyacentes distan 3,4 angstroms a lo largo del eje y se desplazan sucesivamente un ángulo de 36 grados alrededor del eje. Por tanto, la estructura se repite en cada cadena cada 10 bases (360 grados), o a intervalos de 34 angstroms. La información genética está contenida en la secuencia de bases de cada cadena. La secuencia *CATTACTAG* de una cadena se identifica con la secuencia complementaria opuesta, *GTAATGATC*, de la otra cadena.

otros productos farmacológicos. Los microorganismos han jugado desde hace tiempo un papel esencial en las industrias alimentarias, destilerías, laboratorios farmacéuticos y plantas químicas, y un control genético más exacto de sus características ya ha aportado grandes beneficios. Las técnicas de recombinación del ADN no sólo ofrecen un progreso de las prácticas actuales sino que, además, anuncian un nuevo reino de "producción de bacterias" en el que se unen el rápido y controlado crecimiento de microorganismos con la producción de ciertos elementos específicos de los organismos superiores. Entre estas posibilidades se cuenta la producción de insulina, de factores de coagulación de la sangre y de agentes inmunológicos. Valorar la probabilidad de que esto se consiga algún día resulta tan difícil como valorar los riesgos, pero si se pudiera llevar a cabo con éxito proporcionaría sustanciales beneficios para la economía y para la sociedad. De ahí que haya despertado el interés de potenciales inversores.



LA REPLICACION DEL ADN depende de la relación complementaria entre las secuencias de nucleótidos de las dos cadenas de la molécula de ADN. En condiciones químicas adecuadas, los enlaces por puentes de hidrógeno entre las bases se debilitan y las dos cadenas pueden desenrollarse y separarse. En presencia de enzimas apropiados y nucleótidos libres, puede formarse una nueva cadena junto a la parte expuesta de cada cadena vieja desapareada. La secuencia complementaria que se forma al situarse cada nucleótido junto a su opuesto se une después extremo a extremo mediante el concurso de un enzima específico que enlaza los nucleótidos individuales.

Las normas de los NIH no se pronuncian sobre aplicaciones comerciales, sólo estipulan que los experimentos a gran escala (producción de lotes de más de 10 litros) con recombinantes "que se sabe que producen productos nocivos" están prohibidos, a menos que posean una autorización especial. Las normas exigen también informes detallados de los experimentos propuestos con ADN recombinante, previsión que se opone a la defensa de los intereses de los propietarios. Los NIH y los representantes industriales han discutido ampliamente sobre el particular. Además, los portavoces de la industria han prestado declaración ante el Congreso norteamericano. Se sabe que se llevan a cabo algunas investigaciones a nivel industrial y que los representantes de la industria acatan en general las normas preventivas de los NIH, pero no aceptan las limitaciones sobre los derechos de los propietarios y sobre el volumen de producción. Se ha abordado asimismo la política a seguir con las patentes, y el departamento de Comercio de los Estados Unidos tiene ciertas dudas sobre cómo tratarlo. En realidad, las posibles aplicaciones comerciales de las técnicas de recombinación del ADN tienen que debatirse públicamente con la seriedad que merecen y deben situarse en lugar preferente en el marco de las próximas sesiones.

Los problemas de las aplicaciones comerciales van desde las consecuencias inmediatas a otras más amplias y a las de largo plazo. Las técnicas de recombinación del ADN han resucitado el debate sobre "ingeniería genética" y han despertado las viejas disputas sobre la aplicación de la investigación biomédica fundamental a la tecnología, a la calidad de vida y al futuro de la sociedad. La recombinación del ADN se ha sumado, en el cuadro de las "perspectivas pesimistas", al de la fisión nuclear, la superpoblación, el hambre y la escasez de recursos. Se hace más difícil tratar objetivamente estos temas que los relacionados con los posibles daños biológicos, pero surgen abiertamente en los debates públicos en general y en las declaraciones de reputados hombres de ciencia.

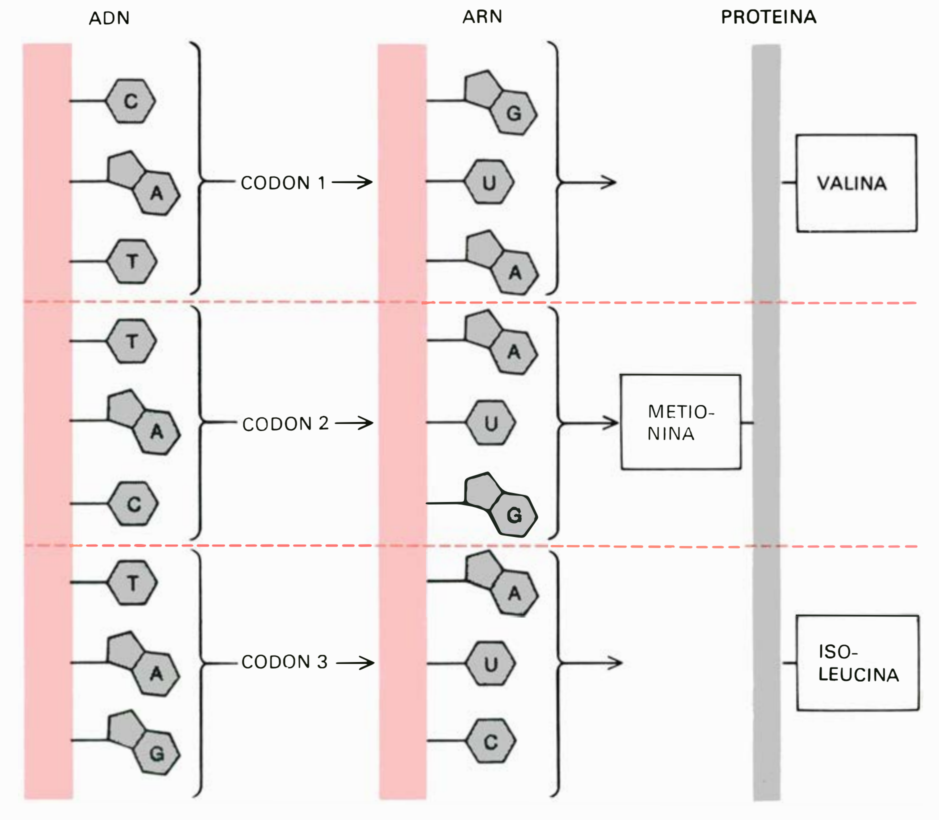
Así, Robert L. Sinsheimer, del Instituto de Tecnología de California, ha planteado insistentemente temas que son en parte prácticos y en parte especulativos o filosóficos. Junto con George Wald, de la Universidad de Harvard, y Erwin H. Chargaff, de la Facultad de Medicina de la Universidad de Columbia, da a entender que todo el trabajo con ADN recombinante para aclarar la complejidad de los sistemas genéticos superiores es una torpeza. El argumento que esgrime no es que esta tentativa no pueda funcionar sino que los riesgos enormes que comporta son innecesarios, ya que podrían utilizarse otras vías que implicaran menor peligro, aunque fueran más lentas. Sinsheimer insiste en la diferencia fundamental que existe entre los organismos procariotas simples, como las bacterias y la complejidad de los organismos eucariotas, entre los que se encuentran los seres humanos. Los procariotas, que son organismos unicelulares, tienen un cromosoma único y relativamente sencillo que queda libre en el medio celular, mientras que las células de los eucariotas tienen un núcleo rodeado de una membrana, núcleo que contiene varios cromosomas mucho más complejos. Los restos paleontológicos sugieren que los procariotas existían en la tierra mil millones de años o más antes que los eucariotas, más complejos, aparecieran en escena. En opinión de Sinsheimer, a lo largo de la evolución de los eucariotas se ha establecido una barrera genética entre ellos y los procariotas, más allá de la cual los eucariotas han desarrollado sus mecanismos de control genético muy complejos. Al transferir estos mecanismos, que son posiblemente la clave del éxito de la evolución y de la enorme diversidad de los eucariotas, a los procariotas, se producirían, según él, incalculables daños evolutivos. Los procariotas podrían adquirir una mayor eficacia biológica, como competidores y como parásitos anulando una vieja estrategia evolutiva.

El argumento de Sinsheimer ha ganado sólo unos cuantos adeptos entre los biólogos, y él mismo afirma que se mueve en el terreno de la especulación. No obstante, su argumento ha atraído la atención pública y es frecuentemente citado como apoyo a los que se oponen a la continuación de las investigaciones sobre ADN recombinante. Bernard D. Davis, microbiólogo de la Facultad de Medicina de Harvard, al examinar el planteamiento anterior, rechaza la idea de la existencia de una barrera genética entre procariotas y eucariotas. Cree que ha habido una extensa y continua oportunidad de intercambio de ADN entre los dos grupos. Señala que las bacterias pue-

den incorporar ADN libre del medio exterior y que *E. coli* estaría expuesta en el intestino humano a este ADN procedente de células humanas muertas. Los microorganismos pueden incorporar ADN de forma similar en el proceso de descomposición de animales muertos. Por tanto, Davis argumenta que la mayoría de los recombinantes probablemente ya se habrían podido formar en el contexto evolutivo natural y no habrían resultado beneficiosos. Razonando análogamente, y a partir de una información extensa sobre las bacterias patógenas, Davis llega a la conclusión de que, bajo las normas existentes de los NIH, la probabilidad de supervivencia en la naturaleza de los recombinantes procariota-eucariota producidos en el laboratorio es muy pequeña.

Esta disparidad de opiniones sobre un tema biológico importante pone de manifiesto la dificultad de valorar las mismas cuestiones más o menos irrelevantes cuando la información crítica es incompleta. La controversia sobre la proporción riesgo-beneficio adquiere incluso una intensidad mayor cuando el argumento implica además delicados juicios de valor. Aquí de nuevo las preocupaciones de los escépticos y contrarios a las técnicas de recombinación del ADN son expuestas sutilmente por Sinsheimer, a través de las siguientes preguntas: "¿Queremos asumir la responsabilidad básica de la vida en este planeta? ¿Queremos desarrollar nuevas formas de vida para nuestros propios objetivos? ¿Podemos tomar en nuestras manos nuestra propia evolución futura?". Ya que estas cuestiones incluyen ideas tales como responsabilidad, misión y control del futuro, implican claramente consideraciones más allá de la ciencia.

Naturalmente, la especie humana ha ido alterando la vida en este planeta desde los comienzos de la cultura humana. Cuando la caza y el cultivo supuso el paso al estadio de la cría de animales y la agricultura, la elección y las intenciones del hombre empezaron a influir en la evolución de las especies seleccionadas. La selección humana al azar fue reemplazada por la cría deliberada de animales y el cultivo de las plantas; el desarrollo posterior de la cultura humana ha ido alterando, a ojos vista, todo el ecosistema. Además, el avance biocultural de la especie humana, basado en parte en la voluntad humana, altera el acervo de genes del hombre y modifica lentamente las especies de manera impredecible. No obstante, la aparición de las técnicas del ADN recombinante ha incrementado obviamente las perspectivas de la ingeniería genética y puesto de manifiesto la importancia de valorar sus implicaciones. ¿Podemos suponer que el éxito en la introducción de ADN recombinante en *E. coli* irá acompañado de un éxito similar al introducirlos en la especie humana? Y si podemos, ¿cuál es el plazo de tiempo para aplicar esta técnica a la especie humana? ¿Es exacto y responsable sugerir que tenemos casi al alcance de la mano el control de "nuestra propia evolución futura"? Estas cuestiones hay que someterlas, evidentemente, a una valoración científica, y deberían ocupar un lugar importante en el nuevo calendario de las líneas de trabajo a realizar.



EL PAPEL DEL ADN EN LA SINTESIS DE PROTEINAS se pone de manifiesto en este diagrama. El mensaje genético contenido en la secuencia de nucleótidos *CATTACTAG*, por ejemplo, está "escrito" en forma de tripletes "codones" *CAT, TAC,* y *TAG.* Cada codón determina, a través de una serie de pasos intermedios en los que interviene una molécula de ácido ribonucleico (ARN), la posición de un aminoácido determinado en una molécula proteica. Por tanto, la secuencia de nucleótidos de una molécula de ADN dada especifica la correspondiente secuencia de aminoácidos de una proteína determinada; cada codón trinucleótido coloca cada uno de los 20 aminoácidos posibles en posiciones sucesivas de la cadena proteica. Ya que la secuencia de aminoácidos establece, a su vez, la estructura y la función de la proteína, la secuencia de nucleótidos del ADN determina implícitamente cada una de las propiedades del organismo. La letra *U* designa el nucleótido uracil pirimidina, que es un componente de la molécula de ARN (ácido ribonucleico).

Sinsheimer se ha introducido en otra área de controversia que afecta no sólo a la comunidad científica sino también a toda la sociedad. Al afirmar que se necesita, probablemente, tiempo para que la capacidad humana "asimile" los nuevos conocimientos genéticos y los pueda aplicar en la naturaleza de forma inteligente, se pregunta si "hay ciertas materias que sería mejor que permanecieran desconocidas, al menos por un tiempo". Esto es una gran herejía en la comunidad científica, cuya premisa fundamental es que el incremento de conocimientos es el promotor y no el esclavo de otros valores. La desestimación del concepto de "conocimiento prohibído" fue parte del período heroico de comienzos de la ciencia moderna, que implicaba además en aquel momento estar dispuesto a hacer frente a la Inquisición y a la estaca. Tocados por la llama nuclear, y debiendo afrontar ahora las más sutiles implicaciones del lenguaje íntimo de la vida, los científicos del siglo XX no temen la estaca sino el juicio de la historia. Chargaff, pionero de las investigaciones que condujeron al desciframiento del lenguaje genético, dice: "Mi generación, o quizá la que me precede, ha sido la primera que ha librado, bajo el liderazgo de las ciencias exactas, una batalla destructiva y colonial contra la naturaleza. El futuro nos maldecirá por ello".

Sinsheimer y Chargaff, junto con varios filósofos, historiadores y sociólogos

de la ciencia, defienden sin embargo que se debe tener en cuenta las posibles consecuencias que derivan del conocer algo, para decidir, conscientemente, la dirección a tomar en la búsqueda de ese conocimiento. Ningún otro tema hirió con más profundidad el corazón de la ciencia moderna. La propia duda expresada por algunos científicos refleja la pregunta que todos se plantean en Estados Unidos sobre los beneficios netos de la ciencia y la tecnología. El análisis coste-beneficio es una preocupación viva y se está aplicando cada vez más a la génesis del propio conocimiento. Es difícil evaluar lo que se puede ganar o perder a partir de un nuevo conocimiento determinado; pero resulta más difícil evaluar los costos de no tenerlo. Este problema subyace,en menor escala, en la controversia sobre el ADN recombinante. El florecimiento de la genética molecular en Estados Unidos ha sido consecuencia directa de una serie de decisiones tomadas después de la segunda guerra mundial que proporcionaron fondos para la investigación biomédica. El objetivo era la conquista de las enfermedades "que matan": cáncer, enfermedades coronarias y embolia. Estas enfermedades siguen siendo relativamente frecuentes, aunque sabemos más sobre ellas y se tratan mejor. Al mismo tiempo, aparte de la investigación subvencionada por el Gobierno Federal, surgió el entusiasmo que llevó al descubrimiento de la doble hélice, el código genético la estructura de las proteínas y el ADN recombinante. Como en un doble juego clásico, el público se pregunta ahora si ha estado comprando salud y bienestar o monstruos quiméricos. ¿Es la genética molecular y toda la tecnología biomédica un aprendiz de brujo? ¿Estamos incrementando, y no disminuyendo, nuestra carga de dolor y ansiedad?

La última cuestión nos lleva a otro tema. Los posibles peligros biológicos y ecológicos pueden surgir inadvertidamente, pero los daños sociológicos pueden ser el producto de una malevolencia deliberada. Los Estados Unidos firmaron un convenio legal internacional en el cual renunciaban a la guerra biológica, abarcando ésta la misma investigación para producir los agentes necesarios. No todos los países han dado ese paso, y una renuncia pública sin una inspección adecuada no puede asegurar que no existan actividades encubiertas. Los contrarios a la investigación sobre el ADN recombinante ven estas técnicas ideales para servir propósitos malévolos, como agentes tanto de una guerra organizada como

LA TECNICA DE RECOMBINACION DEL ADN hace posible, por primera vez, introducir deliberadamente secuencias de nucleótidos procedentes del ADN de una cepa o una especie en el ADN de otra. El ADN del organismo "extraño" se trata primero con enzimas de restricción, que rompen la molécula de doble cadena al azar reconociendo secuencias de nucleótidos determinadas (que normalmente están separadas por miles de pares de bases). El mismo enzima se utiliza después para romper el ADN de un vector apropiado, en este caso un plásmido aislado de una bacteria *E. coli.* Ya que la rotura causada por el enzima no ocurre en el mismo punto en ambas cadenas, dicho tratamiento químico produce una mezcla de segmentos de ADN que tienen extremos de cadena simple complementarios. En condiciones apropiadas, los extremos "pegajosos" de dos secuencias diferentes pueden unirse y formar una única molécula de ADN. Por ejemplo, después de recombinarse el ADN estraño con el ADN del plásmido, el plásmido recupera su forma circular y puede insertarse en una célula huésped apropiada (en este caso *E. coli*) donde puede iniciar nuevamente la replicación, propagando de este modo un número indefinido de copias "clonadas" de la secuencia de nucleótidos insertada, procedente de otro organismo dado.

del sabotaje y del terrorismo. Las técnicas no requieren grandes instalaciones ni instrumentación muy sofisticada. Los que sostienen el punto de vista contrario no lo niegan pero señalan que las técnicas de recombinación del ADN no serían las primeras que tienen, en potencia, aplicaciones torcidas. Los explosivos también las tienen, y la sociedad no los prohíbe de raíz, sino que toma precauciones contra su mal uso.

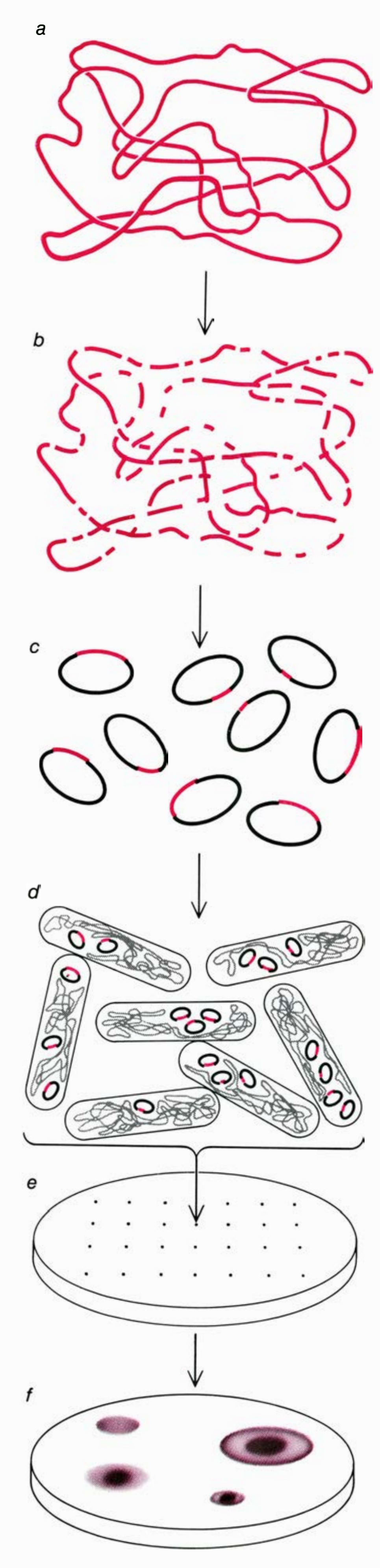
Sin embargo, el tema del posible mal uso de la tecnología del ADN recombinante merece incluirse en el programa del plan de acción a seguir en próximas sesiones, porque pone de relieve la necesidad de un examen a nivel internacional de las implicaciones y el control de las investigaciones y aplicaciones del ADN recombinante. Se podría contestar que los Estados Unidos no están preparados para este análisis público hasta que su propia política al respecto no quede más definida. Pero ya va siendo hora de acometer una reflexión interna sobre cómo se debería presentar en un debate internacional.

Estos son los principales puntos que han sobresalido hasta ahora de las discusiones. No es ninguna lista insignificante. El debate no ha arrancado la pasión en cada esquina, pero llegó a ser lo suficientemente enérgico en Cambridge, el verano pasado, como para repercutir en el continente. Así, una valoración presentada por un grupo de no científicos al Cambridge City Council no difería mucho, en su contenido, de la realizada por un grupo de trabajo del Quality of Life Board, de la ciudad de San Diego, en donde resido. Ambos grupos aceptaban en su comunidad la continuación de la investigación sobre ADN recombinante que exigiera medios de trabajo del tipo *P*3, pero solicitaban una mayor seguridad que la proporcionada por las normas dimanadas de los NIH. Al mismo tiempo, el ministro de Justicia del estado de Nueva York presidió un coloquio público y se llevaron a cabo otros dos más conjuntamente dirigidos por la Asamblea del Cuerpo Legislativo del estado de California. La legislación que regula la investigación sobre el ADN se introdujo más tarde en la Asamblea de California, y todavía se trabaja en ella. El Congreso de los Estados Unidos ha convocado varios debates públicos y se ha aprobado una normativa legal por parte del Senado y por la Cámara de Representantes del Congreso. Estas iniciativas a nivel local, estatal y federal en Estados Unidos apoyan la necesidad de seguir adelante con el plan de acción propuesto en el calendario de las reuniones.

El plan de acción debe considerarse en dos plazos de tiempo por lo menos: a corto plazo (de inmediato) y a largo plazo. Ha ido ganando terreno la idea de que es necesario dar inmediatamente una base legal a la regulación que se encuentra en las normas de los NIH. En particular, la regulación debe extenderse a las actividades no financiadas por organismos federales, especialmente en el sector industrial. De cualquier modo que se haga, no obstante, importa mantener una flexibilidad, ya que los problemas que se han de tratar cambiarán al ir adquiriéndose un mayor conocimiento y experiencia.

Además, dada la complejidad de los aspectos realizados a largo plazo, convendría que la legislación inmediata fuera provisional y limitada en su alcance. Debería crearse asimismo un mecanismo que condujera activamente hacia una política futura más definida. Esto requiere una nueva y amplia valoración de todos los aspectos que han surgido en la investigación del ADN.

La necesidad de una nueva valoración a escala nacional se pone de manifiesto una vez compensado únicamente el resultado de la valoración anterior. Se ha alegado, en primer lugar, que la conferencia de Asilomar de 1975, que establecía el modelo de las normas de los NIH, fue dominada por científicos inmersos en la investigación, y que por tanto no podían tener una perspectiva suficientemente amplia. En segundo lugar, se ha argumentado que la valoración anterior estaba especialmente dirigida a la cuestión de los posibles daños biológicos y olvidaba por completo otras cuestiones de grave importancia. El paso del tiem-



EL EXPERIMENTO "DE PERDIGONADA" es un tipo de experimento de recombinación del ADN en el que el ADN total de un organismo *(a)* se expone a la acción de los enzimas de restricción para obtener muchos fragmentos *(b)*, que se recombinan después con el ADN procedente de un vector apropiado *(c)*; los recombinantes se vuelven a enlazar, al azar, con el ADN de las células hospedantes de *E. coli (d)*. Las células hospedantes de *E. coli* se extienden luego sobre un sustrato nutritivo *(e)*, a fin de que cada célula receptora, que contiene una secuencia de nucleótidos extraña determinada, pueda formar una colonia *(f)*. Si el experimento tiene éxito, el resultado es una "individualización" de todas las secuencias de nucleótidos de un organismo. Según las normas publicadas el año pasado por los National Institutes of Health, de los Estados Unidos, los experimentos de perdigonada se consideran potencialmente más peligrosos que los que interviene ADN purificado y caracterizado, ya que no se sabe si algunos fragmentados de ADN de función desconocida, o no manifiesta, podrían ocasionar graves problemas hoy inesperados.

po ha añadido varios puntos más: que las circunstancias han cambiado al progresar la investigación, que la experiencia ha aumentado y que hay más gente interesada en el tema.

Quien acometa esta nueva revisión nacional debe examinar, ante todo y con detalle, la situación actual, sin omitir el problema de la eficacia real de los mecanismos de control que aportan las normas dimanadas de los NIH. Hay que dedicar una atención especial a los comités locales de salvaguarda de los peligros biológicos, creados según las normas de los NIH. Después de la responsabilidad asignada al investigador principal, tales comités constituyen la única fuente de vigilancia local y sólo ellos pueden fijar las medidas. En su composición y responsabilidad son únicos, pero en cuestión de autoridad y procedimientos sólo están especificados de una manera general por las normas de los NIH. Necesitarán del estímulo y del apoyo de intereses externos para llevar a cabo su importante tarea. Además, no se ha hecho todavía ninguna previsión para el presupuesto del control técnico, formación del personal y chequeos médicos, lo que puede representar un costo considerable. Como cualquier otro seguro, la seguridad contra los posibles daños biológicos puede comprarse. Este costo se debe soportar como un gasto adicional de la investigación, no como una necesidad más a cubrir con el presupuesto existente.

Hay que examinar asimismo el procedimiento real de las secciones de estudio de los NIH, que según las normas deben ser evaluadores independientes de los posibles daños biológicos y del grado de contención que se debe aplicar. Las secciones de estudio ya van muy sobrecargadas con el trabajo de valoración de la calidad científica. Y aún se pide que estos grupos de control, que sólo trabajan a horas, asuman otra difícil función. Si la responsabilidad se tiene que tomar en serio, habría que contar también con otros gastos adicionales.

Para estar sobre aviso desde el principio, es de especial importancia un sistema de control eficaz que siga las direcciones reales de la investigación del ADN recombinante. Las técnicas implicadas son tan ricas en posibilidades (en lo referente a la investigación fundamental o a su aplicación, a los beneficios y los riesgos) que es esencial un "aviso precoz". Para una detección precoz y para una evaluación de los riesgos y de las ventajas, es esencial que se sigan de manera sistemática los caminos que toman los intereses de los investigadores, desde los programas de ayudas económicas y las comunicaciones informales hasta la publicación del trabajo. No es necesario decir que este sistema de control resulta especialmente difícil en la investigación industrial. Por tanto, quizá sería mejor limitar o posponer ciertos avances a los estudios pendientes más detallados y a un mayor conocimiento de los problemas que se han puesto de manifiesto.

Urge asimismo un esfuerzo decidido en pro de una valoración más exacta de los riesgos y de su limitación. Una primera recomendación al equipo que lleve a cabo la nueva valoración sería que se le explicitara en quién recae la responsabilidad de este tipo de trabajo. Dados los diferentes aspectos que deben considerarse en la regulación de la función de los NIH de promover la investigación relacionada con la salud, la función reguladora habrá de recaer sobre alguna de las instituciones implicadas. Por otra parte, habida cuenta de la necesidad de un estudio cuidadoso sobre las implicaciones que se derivan de confiar en los centros existentes o establecer otro nuevo, es mejor que la asignación temporal continúe recayendo sobre los NIH. Esta solución interina, si es que se adopta, debe ir acompañada de fondos especiales para llevarla a cabo.

Las consideraciones sobre los peligros biológicos y la contención física y biológica han gozado necesariamente de una gran prioridad en esta fase temprana de la investigación sobre el ADN recombinante. Sin embargo, muchos observadores informados creen que estos temas perderán importancia al continuar la investigación y aumentar la experiencia. Por tanto, aunque el estusiasmo actual ha conducido a un enfoque racional de los peligros biológicos, que es una parte esencial de cualquier programa de trabajo sobre el ADN recombinante que se haya desarrollado con éxito, tal óptica no agota las exigencias a más largo plazo, e incluso puede distorsionarlas. Temas que puedan haber aparecido directa o indirectamente podrían pasar a la vanguardia de esta larga carrera.

Por ejemplo, al invertir en la investigación sobre genética fundamental, que puede sacar provecho de las técnicas de recombinación del ADN, ¿qué prioridades relativas debería asignarse a las posibles aplicaciones? En el pasado, la estrategia nacional norteamericana en la investigación biomédica ha consistido en invertir directamente en investigación básica, sin objetivos declarados, mientras se invertía también en objetivos concretos, permitiendo que algunos de estos últimos fondos pasaran también a la investigación básica. Así, un investigador que estudiara la interacción entre virus y células, por ejemplo, podía ser alternativa o simultáneamente subvencionado por ayudas para la investigación básica y por ayudas creadas con la intención de promocionar el desarrollo de una terapia eficaz para el cáncer ¿Cuáles deberían ser las prioridades entre las posibles aplicaciones prácticas de la genética molecular? En cuanto a las prioridades entre las líneas competitivas se pueden incluir la síntesis microbiológica de medicamentos, terapias de genes humanos específicos, mejora de la eficacia de la fotosíntesis, fijación de nitrógeno en plantas de cultivo, aumento de la productividad agrícola, etc. En cada una de estas líneas existen riesgos potenciales y distintos tipos de beneficios; resulta poco probable que todas ellas gocen de una subvención inmediata ¿Es necesaria una nueva estrategia de investigación para las líneas de trabajo que se están desarrollando en la actualidad? Si es así, ¿mediante qué procedimiento se debería formular y cómo se habría de llevar a cabo?

Se admite generalmente que existe una continuidad que va desde la investigación básica, a través de la investigación aplicada y del progreso, hasta la aplicación tecnológica. También se admite que esta continuidad no es ni suave ni totalmente predecible y que depende siempre de motivaciones variables y disposiciones institucionales. Las técnicas de recombinación del ADN son el resultado de la investigación fundamental, subvencionada casi completamente por el Gobierno Federal y por las universidades. De momento, al menos, es probable que las técnicas continúen siendo fundamentales en esta área. Sin embargo, tales técnicas podrían también ser útiles aplicadas a finalidades industriales. Dada la naturaleza de la inversión original, y la complejidad de los temas que se han suscitado, ¿deberían permanecer, al menos durante un tiempo, las aplicaciones tecnológicas bajo control federal? Parte de los beneficios obtenidos ¿deberían aplicarse para compensar en parte la inversión original de los recursos federales? ¿Debería servir esta nueva y prometedora tecnología como prototipo para establecer y mantener un fondo de inversión para financiar de modo estable la investigación básica?

Las posibilidades de la ingeniería genética y del control evolutivo ponen de manifiesto los dilemas fundamentales

CONTENCION FISICA	CONTENCION BIOLOGICA (SOLO PARA SISTEMAS HUESPED DE *E. COLI*)		
	*EK*1	*EK*2	*EK*3
*P*1	ADN procedente de procariotas no patógenos que intercambian genes con *E. coli* en la naturaleza ADN de plásmidos o bacteriófagos de células huésped que intercambian genes con *E. coli* en la naturaleza. (Si el genoma del plásmido o del bacteriófago contiene genes nocivos o si el segmento de ADN está purificado y caracterizado en menos de un 99 por ciento, se requieren elevados niveles de contención)		
*P*2	ADN procedente de células embrionarias o de una línea germinal de vertebrados de sangre fría ADN procedente de otros animales de sangre fría y de eucariotas inferiores (excepto insectos mantenidos en el laboratorio durante menos de 10 generaciones) ADN procedente de plantas (excepto plantas que contienen agentes patógenos conocidos o que producen toxinas) ADN procedente de procariotas patógenos de bajo riesgo que intercambian genes con *E. coli* en la naturaleza ADN de orgánulos de eucariotas no primates. (Para el ADN de orgánulos que está purificado en menos de un 99 por ciento se requieren niveles de contención más elevados)	ADN procedente de vertebrados de sangre fría no en estado embrionario ADN procedente de procariotas patógenos, que implican un riesgo moderado, que intercambian genes con *E. coli* en la naturaleza ADN procedente de procariotas no patógenos que no intercambian genes con *E. coli* en la naturaleza ADN procedente de virus de plantas ADN de orgánulos de primates. (Para el ADN de orgánulos que esté purificado en menos de un 99 por ciento se requieren niveles de contención más elevados) ADN procedente de plásmidos o de bacteriófagos de células huéspedes que no intercambian genes con *E. coli* en la naturaleza. (Si existe el riesgo de que el recombinante pueda incrementar la patogeneidad o el potencial ecológico del huésped, se requieren niveles de contención más elevados)	
*P*3	ADN procedente de procariotas no patógenos que no intercambian genes con *E. coli* en la naturaleza ADN procedente de virus de plantas ADN procedente de plásmidos o de bacteriófagos de células huéspedes que no intercambian genes con *E. coli* en la naturaleza. (Si existe el riesgo de que el recombinante pueda incrementar la patogenidad o el potencial ecológico del huésped, se requieren niveles de contención más elevados)	ADN procedente de tejido embrionario de primate o de una linea germinal ADN procedente de otras células de mamíferos ADN de pájaros ADN procedente de células embrionarias, no embrionarias o de líneas germinales de vertebrados (si el vertebrado produce alguna toxina) ADN procedente de procariotas patógenos, que implican un riesgo moderado, que no intercambian genes con *E. coli* en la naturaleza. ADN procedente de virus de animales (si el ADN clonado no contiene genes nocivos)	ADN procedente de tejido no embrionario de primate ADN procedente de virus de animales (si el ADN clonado contiene genes nocivos)
*P*4		ADN procedente de tejido no embrionario de primate ADN procedente de virus de animales (si el ADN clonado contiene genes nocivos)	

EXPERIMENTOS "DE PERDIGONADA" EN LOS QUE SE USA *E. COLI K*-12 O SUS DERIVADOS COMO CELULAS HUESPED, Y PLASMIDOS, BACTERIOFAGOS U OTROS VIRUS COMO VECTORES DE CLONAJE

EXPERIMENTOS EN LOS QUE GENES "EXTRAÑOS" PUROS Y CARACTERIZADOS TRANSPORTADOS POR PLASMIDOS, BACTERIOFAGOS U OTROS VIRUS SON CLONADOS EN LA CEPA *K*-12 DE *E. COLI*

ALGUNOS EJEMPLOS de los requerimientos para la contención física y biológica establecidos por las normas de los NIH, publicadas en el mes de junio de 1976. Las normas, que sustituyeron al aplazamiento parcial que limitó tales investigaciones durante los dos años anteriores, se basaban en los cálculos "pesimistas" de los posibles riesgos asociados a varias clases de experimentos sobre ADN recombinante. Algunos experimentos están prohibidos, como los relacionados con el ADN procedente de organismos patógenos que se sabe que podrían implicar grandes riesgos; otros experimentos, como los relacionados con ADN procedente de organismos que se sabe que intercambian genes con *E. coli* en la naturaleza, requieren solamente la salvaguarda de un trabajo cuidadoso en el laboratorio (nivel *P*1 de contención física) y el uso de la cepa standard de laboratorio *K*-12 de *E. coli* (nivel *EK*1 de contención biológica). Entre estos extremos, las normas de los NIH prescriben, mediante combinaciones adecuadas, un incremento sucesivo de protección física y biológica al aumentar el nivel de riesgo calculado. (En esta tabla la contención aumenta desde la parte superior izquierda hasta la parte inferior derecha). Así, los niveles *P*2, *P*3, y *P*4 de contención física corresponden a un aislamiento mínimo, moderado y máximo, respectivamente. El nivel *EK*2 de contención biológica se refiere al uso de nuevas cepas *K*-12 "deficientes", que conllevan varios defectos genéticos que hacen prácticamente imposible su supervivencia fuera de las condiciones del laboratorio. El nivel *EK*3 se reserva para un sistema huésped-vector de nivel *EK*2 que haya pasado con éxito las pruebas adicionales. Debido a que los medios de trabajo *P*4 son muy limitados y que todavía no hay ningún sistema bacteriano huésped-vector que satisfaga el criterio *EK*1, según los NIH, los experimentos con ADN recombinante que se están realizando en los Estados Unidos con sistemas hospedantes de *E. coli* se limitan, con pocas excepciones, a los expuestos en los recuadros no sombreados. Los experimentos con sistemas hospedantes animal-virus (generalmente sólo el polioma y el virus SV40) exigen el nivel *P*3 de contención física o el *P*4. Los experimentos a realizar con sistemas hospedantes planta-virus vienen regidos por una contención física especial que es análoga a los sistemas *P*1 al *P*4.

originados por las nuevas capacidades adquiridas a través del conocimiento científico. La sociedad ha entrado en una época de intervención, en la que conscientemente se lleva a cabo cada vez más el funcionamiento automático de los procesos naturales, a través de una intervención consciente de la voluntad humana. Muchos de los fenómenos que en tiempos pasados la humanidad sólo podía considerar como un don o un castigo –un acto de Dios o de la naturaleza– son ahora el resultado parcial de la decisión y de la intervención humana. Si el hombre no puede hoy inventar nuevos organismos o crear la vida, probablemente lo hará muy pronto. Si no puede todavía controlar consciente y completamente el comportamiento de amplios ecosistemas, ese poder no está muy lejos del que ya se ha conseguido. La humildad de los individuos se aparta incomprensiblemente del poder abrumador que anteriormente se asignaba a la voluntad divina. Sin embargo, no ha sido la humildad de los individuos la que les ha conferido esta capacidad naciente ni a la que se recurre hoy para controlarla. Ha sido el contacto social de los individuos, actuando a través de instituciones sociales, lo que nos ha llevado a la situación actual y crítica de controlar el destino. Por imperfectas que sean nuestras instituciones sociales, ellas han constituido la plataforma para la época de la intervención.

El reto que afrontamos, reflejado en la estructura maravillosa y rica de la doble hélice, es el de si podemos crear instituciones capaces de transformar los frutos de una época gobernada por la razón en logros de una época regida por la intervención. Hoy se oyen voces que nos urgen no sólo a rehuir la intervención consciente sino también a desconfiar de las aplicaciones y consecuencias de la propia razón y a limitarlas. Quizás sea necesario constatar de nuevo que, después de todo, fue la selección natural la que condujo a la síntesis de la doble hélice con todas sus implicaciones. Entre ellas se hallan el conocimiento y el juicio humanos, a los que les ha tocado el deber de dirigir los procesos y estructuras sociales que puedan afrontar la manipulación de la doble hélice.

El concepto y el control de la doble hélice señala una nueva frontera en el progreso biocultural. En estos momentos se requiere una visión estereoscópica que abarque tanto el "pesimismo creador" como el "optimismo creador". Ninguno de los dos, por sí solo, puede hacer justicia a los profundos conocimientos revelados a los seres humanos recientemente.

La atmósfera de Marte

Su densidad es inferior a una centésima parte de la de la Tierra, pero aún así, es el principal agente que altera la superficie del planeta. Sus vientos y nubes se asemejan a sus homólogos terrestres

Conway B. Leovy

La brillante serie de éxitos obtenidos en las misiones espaciales a Marte, que empezaron en 1964 y acabaron el año pasado al posarse en su superficie con toda normalidad los dos vehículos Viking, han convertido el planeta en algo casi tan familiar como la Luna. A diferencia de ésta, cuya historia parece esencialmente haber concluido hace mil o dos mil millones de años, Marte todavía está evolucionando y cambiando. Como en la Tierra, en Marte, el agente de cambio más eficaz es la atmósfera del planeta, resultante a su vez de la separación de los componentes iniciales del planeta, proceso que comenzó poco después de condensarse a partir de la primitiva nube de polvo y gas que originó el sistema solar hace 4600 millones de años.

Aun cuando, antes de los vuelos espaciales, ya se había cosechado alguna información acerca de la atmósfera de Marte a partir de observaciones telescópicas, tal información era imposible de comprobar y estaba sujeta a una interpretación errónea. Con la información mucho más precisa suministrada por los vehículos espaciales que han pasado cerca de Marte, o descrito órbitas a su alrededor o, finalmente, los que se han posado en él, disponemos ahora de un conocimiento de una segunda atmósfera planetaria lo suficientemente detallado como para comprobar nuestra comprensión de la evolución y de los procesos atmosféricos, comprensión fundamentada antes en el único ejemplo al alcance: la atmósfera de la Tierra. Si prescindimos de su composición sorprendentemente distinta, la atmósfera de Marte se comporta de manera similar a una versión enrarecida de la nuestra. Transporta agua, origina nubes y muestra variaciones diurnas y estacionales de los vientos. En respuesta a los cambios estacionales de la energía recibida del Sol, se producen tempestades de polvo localizadas, cuya intensidad aumenta a veces hasta recubrir la totalidad del planeta. A escala global, tales tempestades de polvo constituyen un fenómeno exclusivamente marciano.

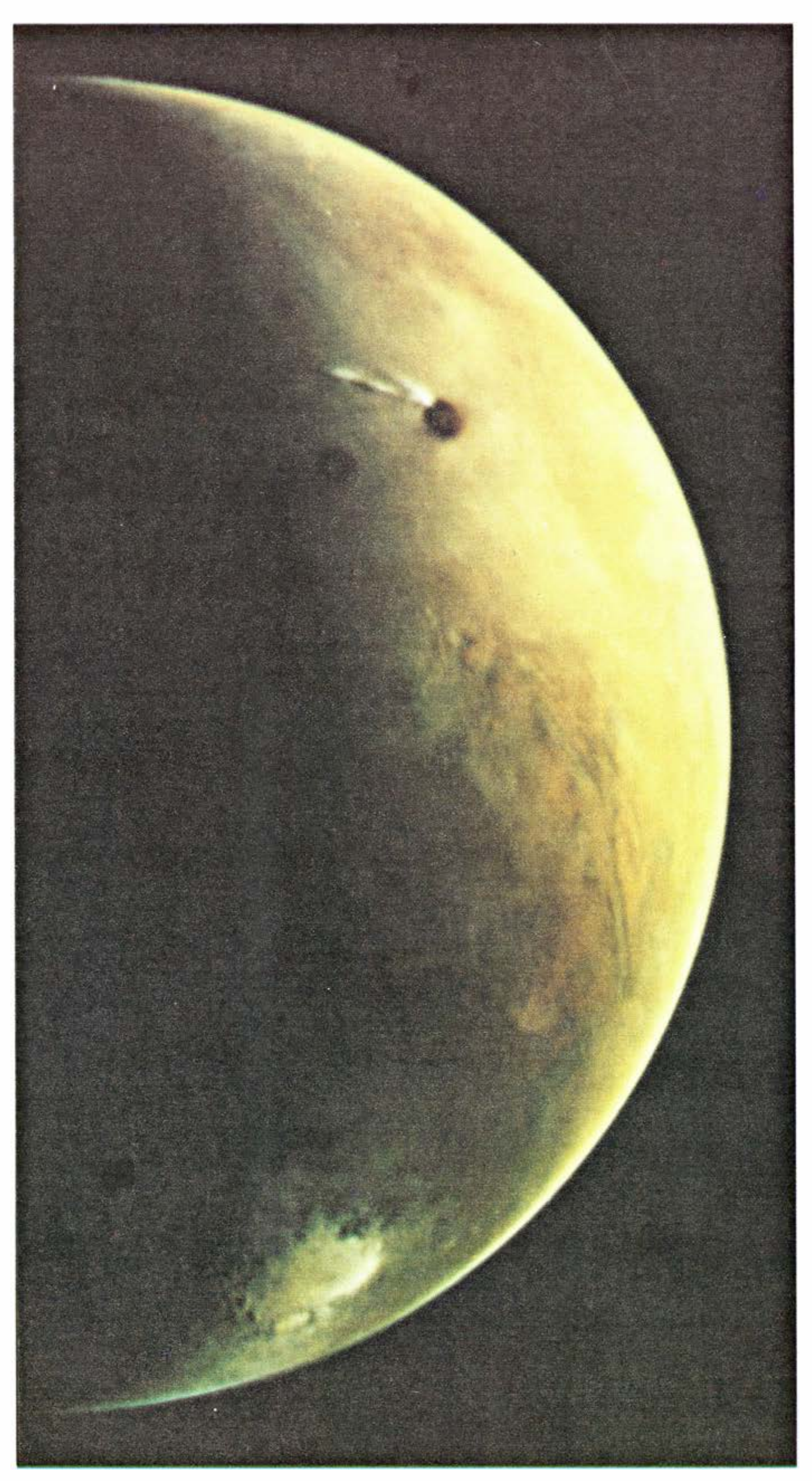
Como resultado de la erosión atmosférica, la primitiva roca cristalina de Marte se ha disgregado en partículas finas, se ha oxidado y se ha combinado químicamente con agua para producir los minerales de color rojizo característico, que tan claramente se aprecian en las fotografías tomadas por los Viking. Los mecanismos de erosión en Marte son claramente diferentes de los que predominan en la Tierra, donde dependen de la presencia de agua líquida. En Marte se encuentran en todas partes pruebas de erosión y sedimentación por viento. Por ejemplo, los dos lugares donde se posaron los Viking parecen haber sido intensamente limados por el viento. En las regiones polares marcianas se han acumulado espesas capas de sedimento depositado por el viento, y en la región que rodea el polo norte se han fotografiado campos de dunas más grandes que ninguno de los existentes en la Tierra.

A consecuencia de la baja presión atmosférica de Marte y la baja temperatura reinante, las únicas formas estables de agua que allí se encuentran son vapor y hielo. Pese a la ausencia de agua líquida en la actualidad, algunas fotografías hechas desde vehículos en órbita demuestran claramente que la superficie del planeta ha experimentado en gran escala erosión debida a líquidos, líquidos que probablemente eran agua. Pudo haber existido en tiempos agua corriente en la superficie como consecuencia de la fusión o vaporización de masas de hielo subyacentes. También es posible que en alguna época el clima fuese mas templado y la atmósfera lo suficientemente densa y húmeda para producir lluvia.

Los atributos más fundamentales de una atmósfera planetaria son su composición y su masa o, lo que es igual, su presión atmosférica en la proximidad del suelo. En época tan reciente como 1963, año anterior al del lanzamiento del *Mariner 4,* que viajó satisfactoriamente a Marte, se creía, a partir de observaciones telescópicas, que la presión atmosférica en la superficie de Marte era de unos 85 milibares, valor algo inferior al 10 por ciento de la presión media en la superficie terrestre, que es de 1013 milibares. También se creía que el gas más abundante en la atmósfera marciana era el nitrógeno. El dióxido de carbono y el agua se habían identificado espectroscópicamente, pero se suponía que eran componentes secundarios porque ninguno de ellos se encontraba en abundancia suficiente para contribuir

NIEBLA HELADA que, a primeras horas de la mañana, cubre de blanco el suelo de las profundas depresiones en la parte occidental de la región de Marte conocida por Coprates Chasma. Finas nubes de hielo se mueven también por encima de los valles y "mesas" circundantes. La repetición diaria de tales nieblas sugiere que el vapor de agua se condensa sobre el suelo por la noche en forma de hielo y se evapora a la salida del sol, volviéndose entonces a congelar conforme asciende en el seno de la fría atmósfera. La fotografía es una composición de tres imágenes en blanco y negro hechas a través de filtros violeta, verde y naranja desde el *Viking 1* en órbita en julio de 1976. Tanto ésta como otras fotografías de los Viking nos han sido suministradas por Geoffrey A. Briggs, del equipo que procesaba las imágenes tomadas por los vehículos orbitales.





MARTE CRECIENTE fotografiado desde el vehículo espacial *Viking 2,* en agosto de 1976, desde una distancia de 450.000 kilómetros. La gran configuración circular situada hacia el sur, la Depresión Argyre, está recubierta de una fina escarcha que se extiende desde el casquete polar de invierno. Las configuraciones lineales más hacia el norte son parte del sistema de cañones llamado Valles Marineris, que se extiende a lo largo de 4000 kilómetros paralelamente al ecuador. Más hacia el oeste, cerca del meridiano que limita la mañana, los dos círculos oscuros son las enormes montañas volcánicas Ascraeus Mons y Pavonis Mons. Extendiéndose hacia el oeste desde Ascraeus Mons hay dos densas líneas nubosas que finalmente desaparecen en el borde de la sombra, a 1000 km de distancia. La sombra de las nubes, visible incluso a esta distancia, indica que las nubes están casi tan altas como el volcán, es decir, más de 20 km sobre el nivel medio de la superficie (más de dos veces la altura del Everest). Las nubes indican que está soplando un fuerte viento del este.

más que en una pequeña fracción a la presión total supuesta de 85 milibares. Aunque el nitrógeno no se puede observar espectroscópicamente, parecía el componente principal más lógico en vista de su abundancia en la atmósfera terrestre.

La estima de la presión en 85 milibares se basaba en la intensidad y polarización de la luz solar reflejada, haciendo uso de ciertas hipótesis para separar los efectos de la luz reflejada por la superficie del planeta de los efectos de la luz difundida o reflejada por la atmósfera. Ahora sabemos que la fracción de la reflexión atribuida a los componentes gaseosos de la atmósfera era exageradamente alta porque se dejó de tener en cuenta la gran cantidad de polvo y calima que más tarde se demostró existía en ella. Aunque la gravedad en la superficie de Marte sólo es el 38 por ciento de la de la Tierra, no hay en principio razón que impida que Marte pudiera retener una atmósfera con una presión casi tan alta como la de la Tierra. El hidrógeno y el helio son los únicos gases cuya energía cinética media en forma de agitación molecular térmica es más que suficiente para permitirles escapar de la fuerza gravitatoria que los sujeta en cualquiera de esos planetas.

Mientras el *Mariner 4* estaba en ruta hacia su destino, un análisis de datos espectroscópicos mejores demostró que la presión en la superficie de Marte tenía que ser mucho menor que 85 milibares. Cuando el *Mariner 4* pasó por la proximidad del planeta en julio de 1965 se pudo establecer con precisión cuánto menor. Desde el punto de vista de los observadores terrestres, el *Mariner 4* desapareció detrás de Marte y reapareció 54 minutos más tarde. Cuando las señales de radio transmitidas por el vehículo pasaron rasando el borde del planeta, fueron refractadas por la atmósfera. La cuantía de la refracción indicaba una presión en superficie de entre cinco y siete milibares. Los nuevos datos espectrográficos habían mostrado la existencia de suficiente dióxido de carbono para producir una presión de ese orden, de modo que el dióxido de carbono pasaba a ser el componente más importante de la atmósfera marciana. El nitrógeno fue relegado a la categoría de componente secundario, si es que en realidad existía.

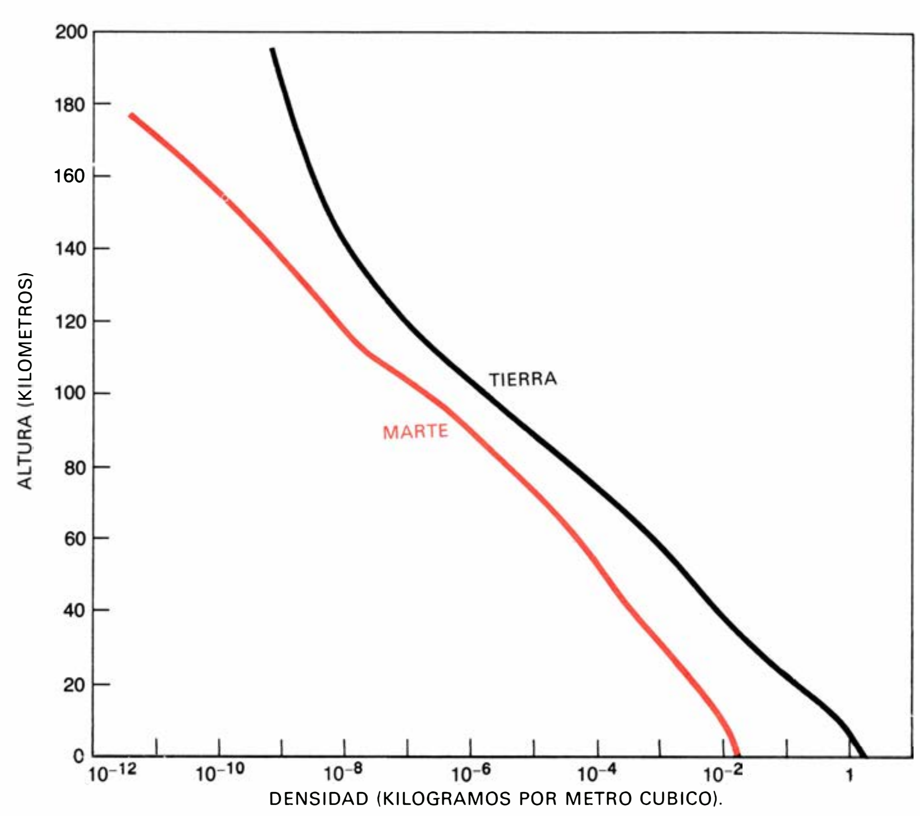
El nitrógeno y los gases nobles (helio, argón, neón, kriptón y xenón) fueron medidos por los dos vehículos Viking durante su descenso a través de la atmósfera marciana y posteriormente en

la superficie del planeta. Los análisis se efectuaron mediante espectrómetros de masas, instrumentos que primero ionizan en el vacío la muestra de gas, y después clasifican los iones resultantes según su masa. Se usaron dos espectrómetros diferentes: uno que operaba en el casi vacío de la alta atmósfera de Marte cuando el vehículo estaba atravesando la región comprendida entre 200 y 120 kilómetros sobre el suelo y otro, proyectado primordialmente para análisis de suelos, en el que el vacío se creaba artificialmente. Aunque no se hicieron medidas entre 120 kilómetros y el suelo, la composición de la atmósfera en la región no explorada se puede deducir con considerable confianza.

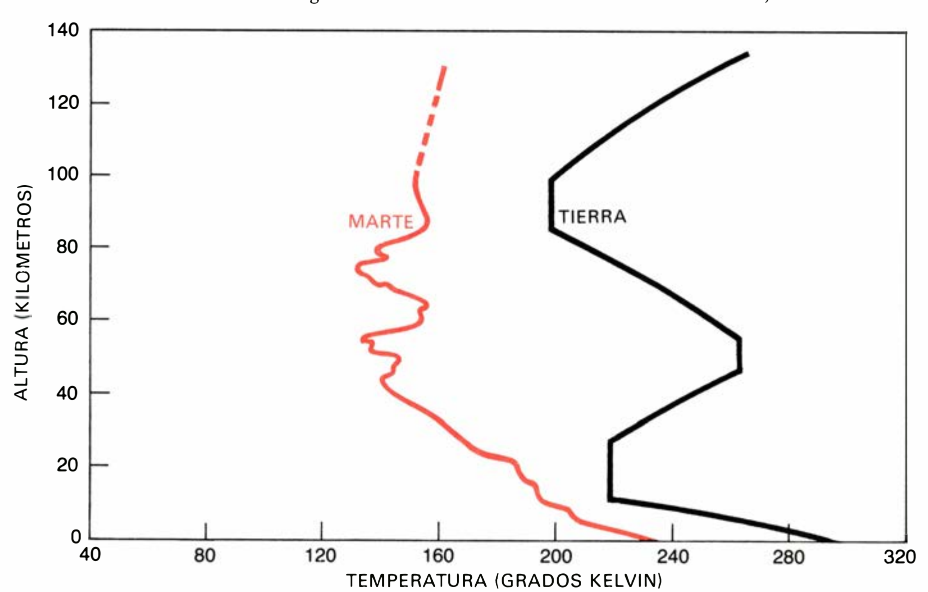
Las medidas en superficie muestran que las moléculas de dióxido de carbono constituyen aproximadamente el 96 por ciento del total. El gas siguiente en abundancia es el nitrógeno, con 2,5 por ciento, seguido por el más abundante de los isótopos del argón, el argón 40, con 1,5 por ciento. En concordancia con medidas anteriores hechas desde la Tierra, existen moléculas de oxígeno en una proporción de sólo 0,1 por ciento. El kriptón y el xenón existen también en cantidades pequeñas, pero determinables, probablemente junto con indicios de neón y helio.

Tanto en Marte como en la Tierra, la atmósfera está constituida por gases liberados del interior del planeta por el volcanismo y otras formas de escape menos violentas. Solamente una parte de los gases liberados, sin embargo, permanece en las dos atmósferas tal y como las observamos hoy. En la Tierra, la mayor parte de las substancias volátiles liberadas están hoy en otros depósitos. El depósito de la mayor parte del vapor de agua liberado es, naturalmente, el océano. La mayor parte del dióxido de carbono se encuentra ahora atrapado en forma de carbonato cálcico en ciertas rocas, como las calizas, que se formaron en el fondo del océano. Otros depósitos de carbono, que originariamente escapó en forma de dióxido de carbono, están representados por yacimientos de carbón, petróleo y esquistos petrolíferos y por carbono disuelto en los océanos y dispersado en la biósfera. La cantidad de carbono almacenada en forma de dióxido de carbono en la atmósfera terrestre es insignificante.

En Marte puede haber también grandes depósitos no atmosféricos de substancias volátiles liberadas. El agua podría estar atrapada en forma de hielo polar o extensamente distribuida en el



LA DENSIDAD DE LA ATMOSFERA MARCIANA es sólo de alrededor de una centésima parte de la de la atmósfera terrestre al nivel del mar. Hasta una altura de unos 100 kilómetros, la densidad de la atmósfera terrestre decrece más rápidamente que la de la atmósfera marciana. Por encima de 100 kilómetros la situación se invierte. El perfil vertical de la densidad de la atmósfera marciana por debajo de 100 kilómetros se ha deducido de un análisis del comportamiento del *Viking 2,* que se posó en el planeta, mientras atravesaba la atmósfera camino de la superficie. El análisis fue realizado por Alvin Seiff y Donn B. Kirk, del Centro de Investigación Ames de la Administración Nacional de Aeronáutica y del Espacio (NASA). Por encima de 100 kilómetros, la densidad atmosférica se dedujo de un análisis de los datos suministrados por el espectrómetro de masas montado en el *Viking 2.* Este análisis lo realizó Alfred O. C. Nier, de Minnesota.



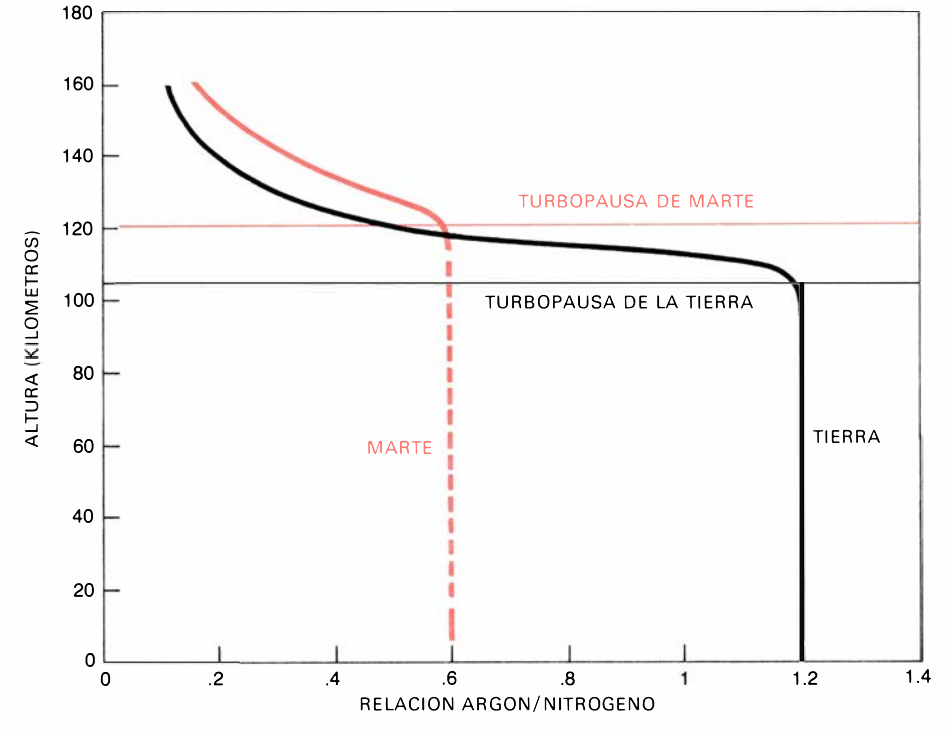
LA TEMPERATURA DE LA ATMOSFERA MARCIANA discrepa marcadamente del régimen en la atmósfera terrestre. La única característica común a ambas atmósferas es la disminución constante de temperatura con la altura hasta unos 10 kilómetros. Por debajo de esa altura, ambas atmósferas se calientan por absorción de energía del suelo, el cual se calienta directamente por los rayos solares. El máximo de temperatura en la atmósfera terrestre, hacia los 50 kilómetros, se debe a la absorción de radiación ultravioleta por el ozono. No hay capa de ozono análoga que absorba energía en la atmósfera marciana. La atmósfera terrestre por encima de 100 kilómetros también se calienta por radiación ultravioleta. Un calentamiento comparable vuelve a estar ausente en la atmósfera de Marte. Las irregularidades en el perfil de temperatura de Marte entre 50 y 100 kilómetros se deben probablemente al enfriamiento por dilatación de capas de gas ascendentes y al calentamiento por compresión de otras que descienden y sustituyen a las anteriores. Tal estratificación puede ser producida por un sistema global de vientos de marea. La temperatura marciana en la vecindad de 130 kilómetros se obtuvo a partir de un experimento Viking de entrada analizado por el profesor Hanson. Las restantes temperaturas proceden de Seiff y Kirk.

subsuelo helado. Se sabe que cierta parte de agua está combinada químicamente en el suelo. El suelo puede retener asimismo grandes cantidades de dióxido de carbono, y es posible que algo de éste exista, helado, en el pequeño casquete polar permanente al sur del planeta. Como la cantidad de materias volátiles almacenadas en tales posibles depósitos es difícil de estimar, se necesitan indicaciones adicionales para determinar cuál ha sido el volumen total de agua y dióxido de carbono que escaparon en forma gaseosa. La abundancia de nitrógeno y gases nobles en la atmósfera marciana puede suministrar tales indicaciones.

Una de las líneas de razonamiento se desarrolla como sigue. Supongamos que la Tierra y Marte fueran originariamente de composición análoga, al menos por lo que respecta al carbono, hidrógeno y gases nobles. Estos no se solidifican ni reaccionan químicamente, y, con la excepción del helio, tampoco se escapan a través de la cima de la atmósfera. Así pues, la relación entre gases nobles en la atmósfera y cantidad total de dióxido de carbono y agua escapados en forma gaseosa a lo largo de la historia del sistema solar debería ser aproximadamente la misma para Marte y la Tierra. Este es el llamado "modelo de analogía terrestre".

Los gases nobles más adecuados para esta comparación son el kriptón y los dos isótopos raros del argón, el argón 38 y el argón 36. El isótopo mas abundante, el argón 40, es un producto de la desintegración radiactiva del potasio 40, y es, por tanto, un indicador más apropiado de la disponibilidad de potasio que del total de materia gaseosa liberada. Las medidas de los Viking en Marte han demostrado que las abundancias relativas de los dos isótopos de argón y sus relaciones al kriptón son semejantes a los valores medidos en la Tierra, lo que concuerda con el modelo de analogía terrestre. Es arriesgado extrapolar el modelo al dióxido de carbono, pero, si se hace, se encuentra que la masa total de dióxido de carbono que ha escapado a la atmósfera en Marte es solamente unas diez veces la cantidad que ahora se encuentra en ella. Se trata de una cantidad sorprendentemente modesta, y sólo con que sea aproximadamente correcta, indica que la cantidad total de gas escapado a la atmósfera en Marte ha sido muchos cientos de veces inferior a la que lo ha hecho en la Tierra.

Una manera completamente diferente de estimar el volumen total de materia gaseosa escapada a la atmósfera en Marte es examinar la abundancia de isótopos de nitrógeno, oxígeno y carbono en la atmósfera para ver en cuánto se han enriquecido los isótopos más pesados de esos elementos con respecto a los más ligeros, que tienen una mayor probabilidad de escapar a través de la cima de la atmósfera. Aunque ni el nitrógeno ni el oxígeno poseen suficiente energía cinética en forma de agitación térmica para escapar de la atmósfera, ambas clases de moléculas pueden escapar lentamente como resultado de su ionización por la radiación ultravioleta en la atmósfera superior. Cuando las moléculas ionizadas y los electrones se recombinan, la energía liberada es más que suficiente para disociar las moléculas en los átomos que las forman. La energía excedente suministra a los átomos de oxígeno y nitrógeno suficiente energía cinética para que puedan ir escapando lentamente. Este mismo proceso no puede actuar en la Tierra debido a que el campo gravitatorio de ésta es mucho más intenso.



LA RELACION DE ARGON A NITROGENO en las atmósferas de Marte y la Tierra decrece paralelamente por encima de unos 120 kilómetros, pero exhibe valores marcadamente diferentes por debajo de las respectivas turbopausas de los dos planetas. La turbopausa marca la zona de transición entre la región inferior de la atmósfera, donde los vientos mantienen bien mezclados los distintos gases componentes, y la región superior, de baja densidad, donde los componentes pueden llegar a separarse por difusión molecular. Por encima de la turbopausa, la relación argón/nitrógeno cae bruscamente a medida que el nitrógeno, el más ligero de los dos gases, aumenta su abundancia relativa antes de su lento escape al espacio. El perfil marciano por encima de 110 kilómetros se ha deducido de experimentos de entrada con los Viking. Los valores en la proximidad de la superficie marciana se han obtenido del espectrómetro de masas montado en el *Viking 2*, en un experimento que ha sido realizado por Klaus Biemann en colaboración con Tobias C. Owen.

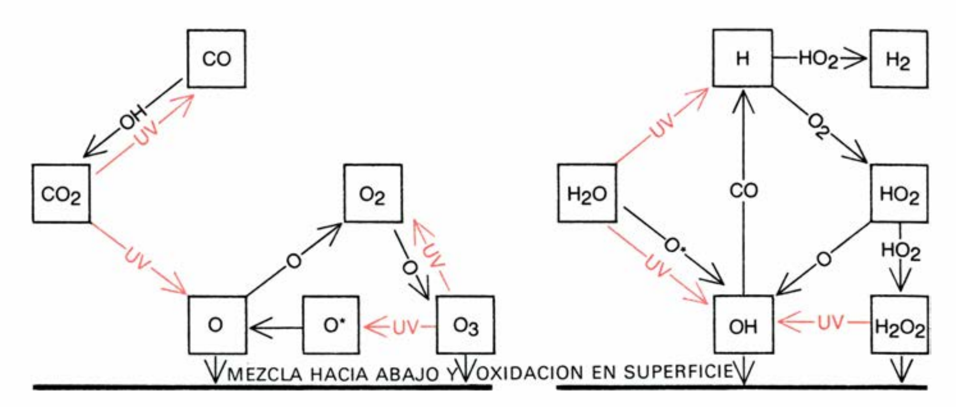
Las medidas de los Viking muestran que las abundancias de los diversos isótopos permanecen constantes hasta unos 120 kilómetros. Por encima de esa altura, las relaciones comienzan a cambiar porque la turbulencia es insuficiente para mantener los gases bien mezclados. En la atmósfera ultraenrarecida que está por encima de 120 kilómetros, las concentraciones de los distintos gases están gobernadas por la difusión molecular, proceso que permite que cada gas se distribuya de acuerdo con su masa molecular, de modo que los gases más ligeros asciendan hacia la cima y finalmente pueden escapar al espacio. La zona que separa la región inferior, bien mezclada, de la atmósfera de la región donde domina la difusión se llama turbopausa. En Marte, la turbopausa está situada hacia unos 120 kilómetros, donde la densidad de la atmósfera es aproximadamente la centésima parte de la densidad en la turbopausa terrestre.

El escape hacia el espacio desde una atmósfera planetaria es posible solamente para moléculas situadas por encima de la turbopausa, porque solamente allí pueden escapar antes de chocar con otras moléculas. Como los isótopos más ligeros son relativamente más abundantes por encima de la turbopausa, escaparán más rápidamente que sus homónimos

más pesados. Si este proceso continuase durante tiempo suficiente, daría lugar a un enriquecimiento selectivo de los isótopos pesados con respecto a los ligeros en todo el seno de la atmósfera. Las medidas de los Viking muestran que la relación del isótopo raro y pesado del nitrógeno, nitrógeno 15, al isótopo común, nitrógeno 14, es aproximadamente un 60 por ciento mayor que la correspondiente relación en la Tierra. El grado de enriquecimiento depende de la cantidad total de nitrógeno liberado, de cuándo fue liberado y de la altura de la turbopausa. Cuanto más alta esté ésta, tanto menor será el enriquecimiento de nitrógeno 14 respecto a nitrógeno 15 en la región desde la cual se puede producir el escape. El enriquecimiento también depende de factores externos tales como el flujo de energía solar ultravioleta, que facilita el proceso de escape suministrando la energía de ionización necesaria. Los modelos de la evolución de la atmósfera marciana basados en la relación de isótopos de nitrógeno sugieren que la liberación de nitrógeno comenzó a principios de la historia del planeta y que el volumen total liberado ha sido grande, quizá cien veces la cantidad actualmente existente en su atmósfera.

Según las medidas de los Viking, no ha habido un enriquecimiento comparable en los isótopos más pesados de carbono y oxígeno. ¿Cómo se puede explicar esto? Una hipótesis es que el carbono y el oxígeno atmosféricos, a diferencia del nitrógeno atmosférico, están en contacto con depósitos grandes, situados en la superficie o bajo ella, de materiales que contienen carbono y oxígeno, de modo que puede haber intercambios atómicos entre la atmósfera y los depósitos. Las cantidades de carbono y oxígeno en la atmósfera pueden representar solamente la cima del témpano. Así, la ausencia de enriquecimiento observable en los isótopos más pesados del carbono y del oxígeno puede ser un argumento en favor de que un gran volumen de gases liberados haya creado un considerable depósito de dióxido de carbono. Parece razonable concluir que la liberación de gases en Marte, aunque bastante menos eficaz que en la Tierra, produjo cantidades de dióxido de carbono y vapor de agua mucho mayores que las ahora halladas en su atmósfera.

A diferencia del nitrógeno y los gases nobles, el vapor de agua es un componente muy variable de la atmósfera marciana. Las medidas realizadas desde la Tierra habían mostrado una variación estacional y sugerido una variación diurna. Los espectrómetros situados a bordo de los Viking en órbita marciana han demostrado ahora que la cantidad de vapor de agua varía marcadamente con la latitud. Aunque la variación vertical del vapor de agua es difícil de establecer espectroscópicamente, la masa total de agua contenida en una columna de sección unidad extendida desde el suelo a la cima de la atmósfera se calcula con facilidad. Convirtiendo dicha masa de vapor de agua en su equivalente líquido, se obtiene el número de "centímetros precipitables", es decir, la altura que alcanzaría el agua si precipitase.



REACCIONES QUIMICAS ATMOSFERICAS se inician cuando el dióxido de carbono (CO_2) y el vapor de agua (H_2O) son disociados por la radiación ultravioleta (UV). En el ciclo del oxígeno *(izquierda)*, el CO_2 se disocia en monóxido de carbono (CO) y oxígeno atómico (O). El último se recombina para formar O_2 y ozono (O_3). Tanto el O como el O_3 contribuyen a la oxidación de rocas y suelo. En el ciclo del hidrógeno *(derecha)*, el H_2O se disocia en hidrógeno atómico (H) y el radical hidroxilo (OH). La reacción de H con el O_2 produce el radical perhidroxilo (HO_2) que puede reaccionar consigo mismo para producir peróxido de hidrógeno (H_2O_2). Tanto el H_2O_2 como el OH puede contribuir a la oxidación de la superficie. Los dos ciclos interaccionan principalmente a través de las reacciones de CO con OH, O_2 con H y O con H_2O. El oxígeno atómico excitado (O*), formado por la disociación ultravioleta del O_3, puede también romper la molécula de H_2O en radicales OH. En el ciclo del hidrógeno se forma también hidrógeno molecular (H_2). Tanto el H_2 como el O_2 son lentamente eliminados de la atmósfera por mezcla hacia arriba y ulterior escape. Los aspectos cuantitativos de los ciclos aquí representados han sido resueltos por Michael B. McElroy, de la Universidad de Harvard, y Thomas M. Donahue, de la de Michigan, e independientemente por Truman D. Parkinson y D. M. Hunten, de Kitt Peak.

Los valores medidos hasta ahora por los Viking en órbita varían desde menos de 10^{-4} centímetros precipitables en altas latitudes meridionales, hacia mitad del invierno, hasta 10^{-2} centímetros precipitables en altas latitudes boreales, hacia mitad del verano. La atmósfera terrestre contiene normalmente de dos a tres centímetros precipitables, de modo que la primera impresión es de que la atmósfera marciana es extremadamente seca. Sin embargo, una comparación más ajustada a la realidad es la hecha con la cantidad total de agua contenida en la atmósfera terrestre por encima del nivel al cual su presión iguala a la presión en la superficie de Marte; o sea, siete milibares. Con este criterio, la atmósfera marciana es muy húmeda: incluso el valor más bajo medido en Marte es varias veces mayor que el contenido de agua de la atmósfera terrestre. Una medida más apropiada es la humedad relativa, o grado de saturación de la atmósfera. Por término medio, la humedad relativa en la atmósfera marciana es tan grande que la abundancia de vapor de agua parece estar controlada en gran parte por la saturación. La atmósfera está casi tan húmeda como puede estarlo a la temperatura atmosférica reinante.

Sin embargo, para mantener la variación observada en el contenido de vapor de agua con la latitud se necesitan, claramente, otros factores distintos de la saturación. Uno importante es la distribución de fuentes y sumideros de vapor de agua. A finales de verano hay en cada polo un pequeño casquete residual de hielo y, al menos en el polo norte, se sabe que el casquete es realmente de agua congelada. La sublimación del hielo bajo el sol estival constituiría una abundante fuente de vapor de agua. No es sorprendente, pues, que las máximas cantidades de vapor de agua observadas por los Viking se encuentren alrededor de los bordes del casquete polar norte. Curiosamente, sin embargo, el *Mariner 9* en órbita no detectó un aumento similar alrededor del casquete residual del polo sur hacia mitad del verano. Si los Viking orbitales logran recoger datos durante un ciclo anual completo, el misterio puede quedar aclarado.

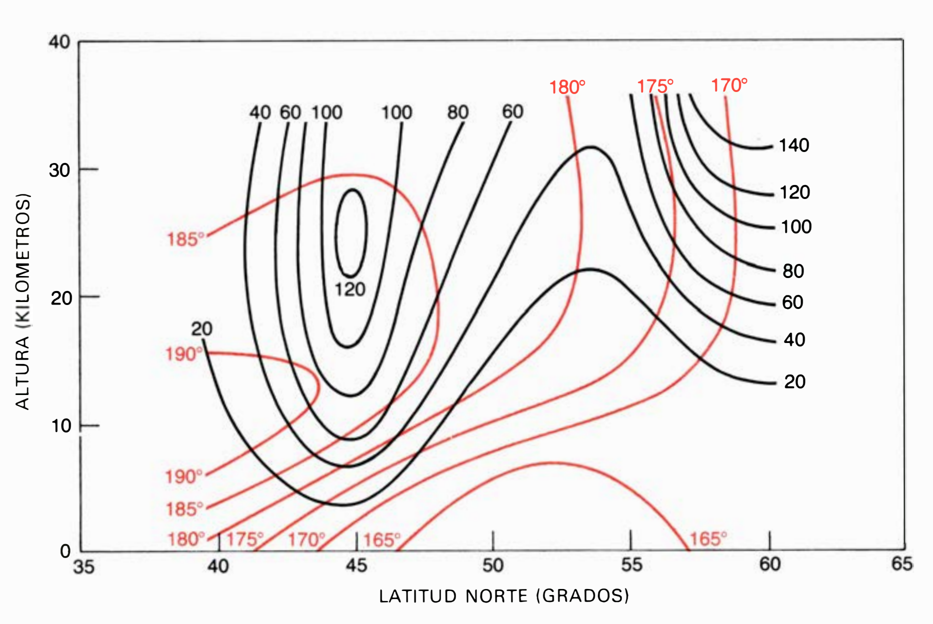
El vapor de agua es el ingrediente clave en la química de la atmósfera marciana. A diferencia de la Tierra, que está protegida de la radiación solar ultravioleta por una densa capa de ozono,

Marte queda expuesto a esta radiación de alta energía hasta su misma superficie. Como resultado, en toda la atmósfera existe disociación del dióxido de carbono en monóxido de carbono (CO) y oxígeno atómico (O), y del vapor de agua en hidrógeno atómico (H) y radical hidroxilo (OH). El H y el OH son extremadamente reactivos. Por ejemplo, pueden catalizar la recombinación del CO y O para volver a formar dióxido de carbono, lo que ayuda a explicar por qué la atmósfera está principalmente compuesta de dióxido de carbono. El hidrógeno atómico y el oxígeno molecular se combinan mediante una serie de reacciones para dar peróxido de hidrógeno (H_2O_2), poderoso oxidante que puede tener una importante influencia en la química del suelo.

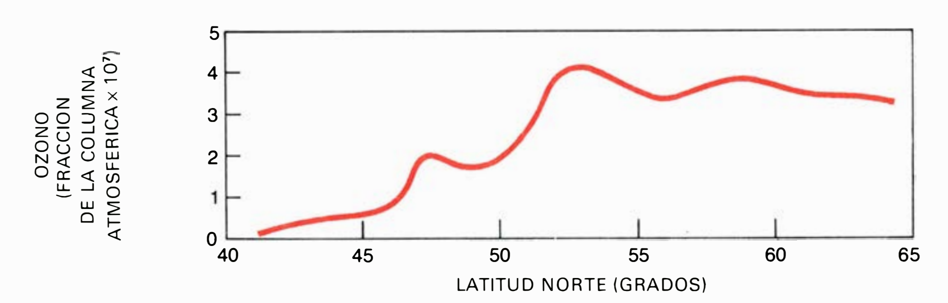
Por disociación del vapor de agua se produce una pequeña cantidad de hidrógeno molecular (H_2). Como es muy estable, el H_2 se halla bien mezclado por debajo de la turbopausa. Por encima, se disocia en hidrógeno atómico, que escapa lentamente. Extrapolando a la totalidad de la vida de Marte (4600 millones de años), la velocidad de escape actualmente estimada puede dar cuenta de la pérdida de una capa de agua líquida de menos de un metro de altura, cantidad que, probablemente, es muy inferior a la cantidad total de agua liberada desde el interior del planeta.

Aunque las cuatro especies que contienen hidrógeno, H, H_2, OH y H_2O_2, son químicamente importantes, están por debajo del umbral de detección que tienen los instrumentos de los Viking. Los procesos químicos que he descrito se deducen de la experiencia terrestre. Los espectrómetros ultravioleta de los vehículos espaciales Mariner, sin embargo, han detectado el hidrógeno atómico que se escapa de la atmósfera superior de Marte. Estas observaciones prestan apoyo a la supuesta disociación del hidrógeno molecular por encima de la turbopausa.

La distribución de ozono (O_3) observada en la atmósfera de Marte proporciona ulterior confirmación de los hipotéticos procesos químicos. Aunque el ozono es mucho menos abundante que en la atmósfera terrestre, hay concentraciones de él fácilmente detectables sobre las regiones polares marcianas en invierno. En latitudes más bajas y más cálidas no suele podérsele detectar. El ozono se produce por combinación de O y O_2, que se han originado por descomposición del dióxido de carbono. El ozono y su precursor, el oxígeno atómico, son destruidos por reacciones catalíticas que dependen del H y OH. Así pues, la concentración de ozono será baja dondequiera que haya buena provisión de H y OH, lo que sucede donde la materia prima, vapor de agua, sea abundante. Como el vapor de agua está virtualmente ausente alrededor del polo en invierno, el ozono se encuentra allí en abundancia. La limitación que la disociación del vapor de agua impone a la concentración de ozono desempeña un papel análogo en la atmósfera terrestre. Por encima de una altitud de aproximadamente 40 kilómetros, la concentración de ozono está limitada por las mismas reacciones catalíticas con H y OH. De aquí que la atmósfera marciana se parezca a la estratosfera terrestre tanto en su química como en su presión.



TEMPERATURA DE LA ATMOSFERA MARCIANA EN INVIERNO *(en color)*, **representada sobre un plano meridiano, deducida de las medidas en el infrarrojo hechas por el** ***Mariner 9*****. Las temperaturas, dadas en grados Kelvin, pueden servir para deducir la distribución de la componente este-oeste de los vientos** *(en negro)*, **dadas en metros por segundo. (Cien metros por segundo equivalen a 360 kilómetros por hora.) Las velocidades del viento aquí presentadas han de sumarse a las del viento en superficie, que son típicamente débiles en invierno. El experimento en infrarrojo del** ***Mariner 9*****, del que se han deducido las temperaturas, se debe al investigador Rudolf A. Hanel, B. J. Conrath y colaboradores del Centro de Vuelo Espacial Goddard, de la NASA.**



DISTRIBUCION DE OZONO EN INVIERNO, representada sobre el mismo plano usado para temperatura y viento en la ilustración superior. El eje vertical presenta el número de moléculas de ozono como fracción del número total de moléculas en una columna vertical de atmósfera. El ozono aumenta hacia el polo, donde las temperaturas y la concentración de vapor de agua son bajas. Los valores fueron obtenidos por Charles A. Barth y sus colaboradores, pertenecientes a la Universidad estatal de Colorado, a partir de los datos en el ultravioleta del ***Mariner 9*****.**

Como la atmósfera marciana está próxima a la saturación, podría esperarse que hubiera gran abundancia de nubes de hielo, como en realidad sucede, al menos durante el verano del hemisferio norte que los Viking han examinado hasta ahora. Como también era de esperar, las nubes de hielo son delgadas y con bordes difusos. No hay nada parecido a los cúmulos densos y de bordes bien definidos que conocemos en la Tierra. Las nubes marcianas son de cuatro tipos generales: convectivas, ondulatorias, orográficas y nieblas.

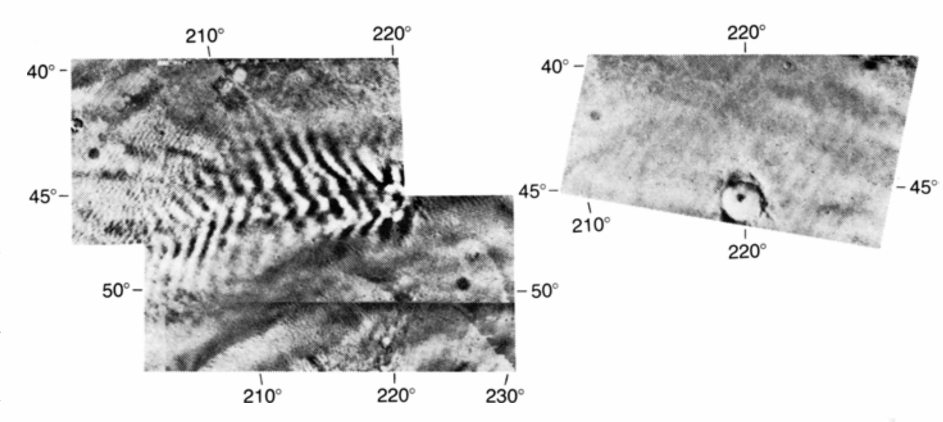
Las nubes convectivas se forman cuando el gas cercano a la superficie se calienta durante el día, asciende y se enfría al dilatarse. Cuando la temperatura desciende hasta el punto de saturación, se forman nubes en forma de borla, casi todas del mismo tamaño e igualmente espaciadas. Tales distribuciones nubo-

sas son frecuentes hacia el mediodía sobre las altiplanicies ecuatoriales. El espacio entre las nubes y sus alturas, estimadas a partir de la situación de sus sombras, sugiere que existe una activa convección en una capa atmosférica de espesor entre cinco y ocho kilómetros, que es algo mayor que el espesor de la capa calentada durante el día sobre las áreas continentales en la Tierra. La poderosa agitación diaria causada por la convección garantiza que todos los componentes atmosféricos de la capa puedan interaccionar eficazmente con la superficie marciana.

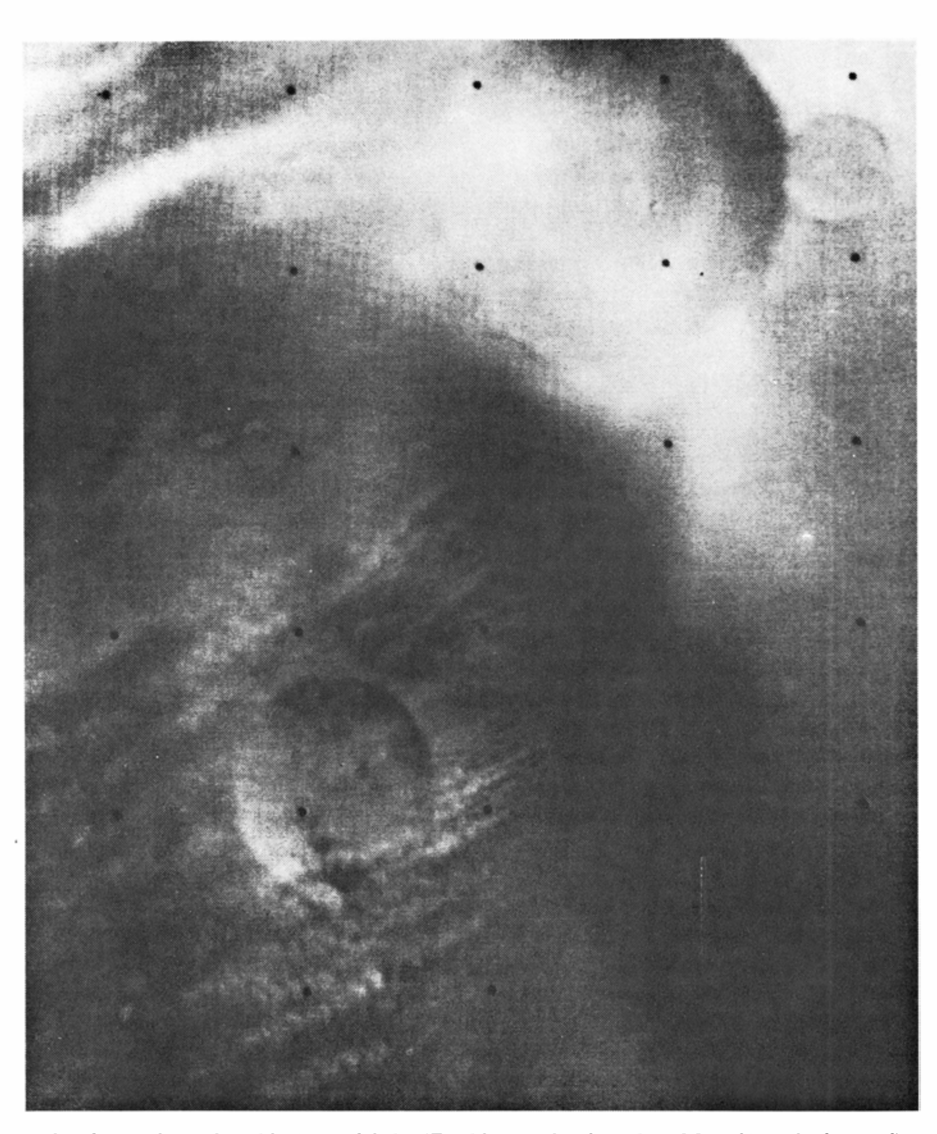
La mejor fotografía de nubes de ondulatoria fue realizada por el *Mariner 9,* que examinó el hemisferio norte en invierno. Las nubes de ondulatoria se forman cuando un viento fuerte sopla constantemente tramontando un obstáculo, tal como una cadena montañosa de suficiente altura. A sotavento del obstáculo se forman trenes de ondas atmosféricas análogas a las olas en el agua. Si las condiciones de temperatura y humedad son las adecuadas, las crestas de las ondas quedan identificadas por nubes que parecen ser estacionarias.

Las nubes de ondulatoria de Marte muestran que, en invierno, la meteorología de las latitudes medias presenten una sorprendente semejanza con la de las regiones análogas de la Tierra. Los vientos dominantes son del oeste a todos los niveles hasta una considerable altura, y los vientos más fuertes, como las corrientes en chorro terrestres, se hallan a varios kilómetros por encima de la superficie. Si se combina la información sobre temperaturas, obtenida por exploración remota, con las fotografías de nubes de ondulatoria, se puede reconstruir un cuadro tridimensional de la distribución del viento en ciertos días. Es típico que el viento sople del oeste con velocidades de 10 a 20 metros por segundo cerca de la superficie, alcanzando más de 100 metros por segundo (360 kilómetros por hora) en la vecindad de la altura de la corriente de chorro, por encima de los 10 kilómetros. Estas velocidades son aproximadamente el doble de lo que se encuentra en la Tierra a las correspondientes alturas.

Las nubes orográficas se forman cuando el gas atmosférico se ve obligado a ascender lenta y constantemente al moverse ladera arriba en una región montañosa. Tales nubes son corrientes sobre las altas regiones volcánicas de Marte. En el verano del hemisferio norte, cuando la atmósfera está húmeda, grandes áreas quedan cubiertas por



NUBES DE ONDULATORIA EN LA ATMOSFERA DE MARTE, hacia mitad de invierno, claramente visibles en el mosaico de tres fotografías de la izquierda, hechas por el *Mariner 9* en 1972. El largo tren de ondas casi oscurece el amplio cráter sobre el cual se han formado las nubes. Las nubes ondulatorias individuales están separadas unos 20 kilómetros, aproximadamente la mitad del diámetro del cráter. El cráter libre de nubes puede verse hacia el borde inferior de la fotografía única de la derecha, que fue tomada exactamente un día antes. Las longitudes y latitudes se indican a lo largo de los bordes de los mosaicos. A fin de impedir que se invierta el relieve aparente del cráter y los otros accidentes en superficie, las fotografías se reproducen con el norte en la parte de abajo, de modo que el sol ilumine los accidentes geográficos desde arriba.



NUBES DE ONDULATORIA A SOTAVENTO DE UN CRATER MARCIANO, fotografiadas por el *Mariner 9* en 1972. La ancha y difusa nube ondulatoria que aparece en la parte superior derecha es análoga a las nubes visibles sobre el cráter de 40 kilómetros en los mosaicos de baja resolución en la parte superior de la página. Las nubes son probablemente de hielo. El cráter que origina estas nubes es solamente de unos siete kilómetros de diámetro. Hacia la parte inferior izquierda y hacia abajo del cráter hay calles de nubes cuyos marcados bordes y estructura en pequeña escala sugiere que están constituidas no de agua, sino de dióxido de carbono sólido (hielo seco), que se puede congelar rápidamente en la atmósfera sobre las regiones polares en invierno.

nubes delgadas bastante uniformes. Durante días y días se repite su proceso de crecimiento y disipación con notable exactitud, indicando la regularidad del tiempo meteorológico de un día para otro.

Una de las más sorprendentes observaciones atmosféricas hechas por los vehículos orbitales Viking fue la de nieblas matinales junto al suelo en varias regiones de tierras bajas. Las nieblas se producen evidentemente cuando la escarcha del suelo se convierte en vapor al ser calentada por el sol matinal. Como las nieblas se ven día tras día en los mismos lugares, parece probable que el vapor de agua atmosférico forme cada noche una nueva capa de escarcha sobre la superficie.

El agua no es la única substancia capaz de formar nubes en Marte. En las regiones polares en invierno y a grandes alturas, la temperatura puede descender lo suficiente para que se condense el dióxido de carbono, creando nubes de hielo seco. Como la condensación y sublimación del dióxido de carbono pueden ser rápidas, las nubes de hielo seco, al contrario que las de hielo verdadero, pueden tener los bordes bien definidos. Algunas nubes brillantes y de bordes muy marcados, fotografiadas por los vehículos espaciales Viking y Mariner, se sospecha sean de hielo seco. Es también posible que durante el invierno puedan producirse tempestades de hielo seco, que ayudarían a crear los casquetes polares estacionales de hielo seco. La mayor parte del hielo seco que constituye esos casquetes, sin embargo, se deposita probablemente cuando el gas dióxido de carbono se pone en contacto directo con la frígida superficie de Marte y se condensa.

Alrededor del 20 por ciento del dióxido de carbono en la atmósfera marciana describe un ciclo entre la atmósfera y el casquete polar de invierno en cada estación, causando la correspondiente variación de la presión atmosférica. La variación de presión, que se aprecia en toda la atmósfera y no solamente en la región del casquete polar, ha sido medida por los barómetros de los dos Viking que se posaron. La magnitud de los cambios de presión concuerda notablemente bien con los modelos teóricos, anteriores a los Viking, de intercambio de dióxido de carbono entre los casquetes polares y la atmósfera. Las medidas muestran que ningún otro gran depósito de dióxido de carbono, por ejemplo el dióxido de carbono adsorbido en las partículas del suelo marciano, lo intercambia estacionalmente con la atmósfera. Puede existir, sin embargo, un gran depósito de dióxido de carbono adsorbido que sea capaz de intercambiarlo con la atmósfera a lo largo de un período de tiempo mucho más grande.

Los vientos de Marte han sido durante mucho tiempo objeto de especulación. En años recientes se han hecho serios esfuerzos para desarrollar modelos teóricos que los expliquen usando técnicas análogas a las empleadas para la predicción del tiempo y clima terrestre. Dichas técnicas deben ser aplicables, ya que ambos planetas presentan la misma sucesión de estaciones (porque sus ejes polares están inclinados casi idénticamente) y tienen la misma velocidad de rotación (24 horas 37 minutos para Marte). La velocidad de rotación es importante porque determina la magnitud de la desviación de masas móviles de atmósfera. La desviación, o fuerza de Coriolis, es la responsable de la rotación de los vientos en sentido antihorario en torno a una zona de bajas presiones en el hemisferio norte y del correspondiente giro en sentido horario en el hemisferio sur.

El esquema que resulta de las investigaciones teóricas es el de una distribución planetaria de vientos con dos regímenes principales. Uno es un régimen de invierno en latitudes medias similar al que existe en la Tierra, con vientos dominantes del oeste, corrientes en chorro a gran altura y perturbaciones o borrascas móviles. El otro es un régimen de verano ecuatorial, libre de perturbaciones viajeras, de hecho libre de todas las variaciones diarias excepto las graduales asociadas a la lenta marcha de las estaciones. Bajo el segundo régimen, la principal influencia sobre la distribución del viento es la variación diurna del calentamiento solar y su interacción con la topografía local. Así, el gas recalentado asciende a lo largo de las laderas de la región durante el día y baja de nuevo durante la noche. Como existen laderas pendientes en casi cualquier parte de Marte, los vientos de pendiente, desviados por la rotación planetaria, deben estar muy generalizados.

Además, el ciclo diario de calentamiento debe crear mareas atmosféricas u oscilaciones en el régimen de viento extendidas a todo el planeta. Como Marte es un planeta de desiertos, las mareas deben ser más acusadas que en la Tierra. Cerca del suelo, los vientos de marea son

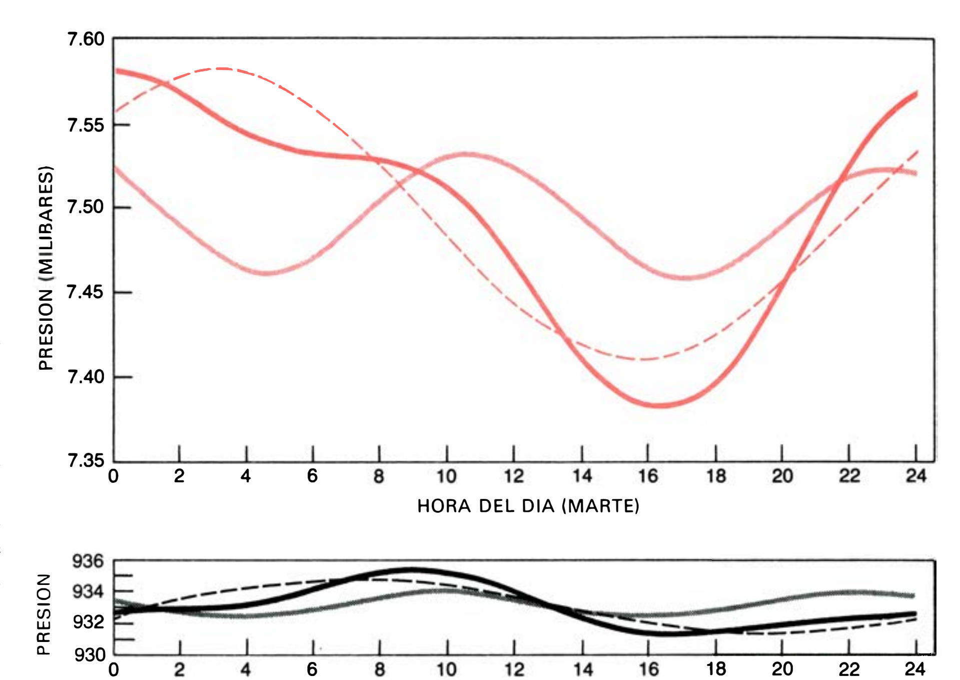
VARIACIONES DIARIAS DE LA TEMPERATURA ATMOSFERICA en el lugar donde se posó el *Viking 1 (en color)*. Son cualitativamente análogas a las del Lago China, California, un lugar desértico *(en negro)*. En ambos casos la temperatura alcanza un mínimo hacia la salida del sol y llega a su máximo unas 10 horas después. El intervalo diario de temperatura, sin embargo, es en Marte unas tres veces mayor que en la Tierra. En el lugar de toma del Viking, el intervalo es de 55 grados, desde 187 a 242 grados Kelvin (-86 a -31 grados Celsius). En el Lago China el intervalo es de 18 grados, desde 292 a 310 grados Kelvin (que corresponde a 19 a 37 Celsius).

flojos: en la Tierra alcanzan a lo más un metro por segundo. En Marte y durante el verano del hemisferio norte deben alcanzar una velocidad entre cinco y diez veces mayor. Las velocidades aumentan bruscamente, sin embargo, con la altitud. En la Tierra alcanzan unos 50 metros por segundo entre alturas de 80 y 100 kilómetros. Aun cuando no se han medido las velocidades correspondientes en Marte, se puede deducir de la altura de la turbopausa que son más grandes. Los vientos de marea son agentes eficaces para agitar la atmósfera superior y, probablemente debido a ello, controlan la altura de la turbopausa en ambos planetas. Como la turbopausa marciana está a un nivel de densidad más bajo que la terrestre, como consecuencia de una agitación más eficaz por los vientos, parece probable que las mareas atmosféricas sean más intensas en Marte que en la Tierra.

¿Apoyan las medidas de los Viking de algún otro modo esta conjetura? Quizá la mejor comprobación de la intensidad de las mareas atmosféricas en Marte la suministren las lecturas barométricas de los dos puntos donde se posaron los Viking. Los instrumentos muestran grandes variaciones diurnas de la presión atmosférica, indicación clara de intensas mareas atmosféricas. Las medidas de los perfiles verticales de temperatura realizadas por los vehículos al penetrar en la atmósfera suministraron una comprobación adicional. A los niveles más altos, donde los vientos de marea han de ser fuertes, deberían observarse grandes variaciones de temperatura originadas por capas atmosféricas alternativas, calentadas unas por compresión y enfriadas otras por expansión. La teoría predice que tales variaciones deben darse con distancias de unos 20 kilómetros y que su amplitud debe aumentar con la altura.

Los perfiles de temperatura registrados durante la entrada de los Viking muestran precisamente variaciones de temperatura con esas propiedades. Parece, pues, que en Marte, como en la Tierra, la altura de la turbopausa viene regida por la intensidad de las mareas atmosféricas. La turbopausa, a su vez, actúa como una especie de válvula que regula el escape de los gases ligeros. Como la turbopausa marciana está relativamente alta, el volumen de gases ligeros disponible para su escape es correspondientemente bajo, y a ello se debe en parte el que el actual ritmo de escape del hidrógeno en Marte sea tan bajo.

Las grandes tempestades de polvo que se producen en Marte se observaron mucho antes de que comenzara la era de la exploración espacial. A causa de la baja densidad de la atmósfera marciana, para producir tempestades de polvo en Marte se necesitan vientos mucho más fuertes que los requeridos en la Tierra. Una velocidad del viento tan baja como es seis o siete metros por segundo (20 a 25 kilómetros por hora) puede comenzar a levantar granos de arena en muchos desiertos terrestres. Experiencias de laboratorio han demostrado que en Marte se requieren vientos de al menos 30 a 60 metros por segundo (110 a 215 kilómetros por hora). Como en Marte no hay lluvia que limpie el aire, pequeñas partículas de polvo pueden permanecer suspendidas durante semanas o meses.



LAS VARIACIONES DIARIAS DE LA PRESION ATMOSFERICA, en relación a la presión total, son unas cinco veces mayores en el lugar de toma del *Viking 1* que en el Lago China en la Tierra. La curva continua en color, en la parte alta, muestra que en el curso de un día marciano la presión atmosférica varía 0,2 milibares, que es aproximadamente el 2,5 por ciento de la presión total media de 7,5 milibares. La curva se puede descomponer en un armónico diurno *(curva de trazos)* y un armónico semidiurno *(curva en color pálido)*. La curva negra de abajo es la variación diaria de presión en el Lago China, que ha sido igualmente descompuesta en dos armónicos. Para igualar la variación relativa diaria en Marte, un barómetro en la Tierra debería recorrer un intervalo de unos 23 milibares en el curso del día. El intervalo real en el Lago China es solamente de 4,5 milibares. Las grandes fluctuaciones diarias en Marte indican que las mareas atmosféricas y los sistemas de vientos que las acompañan son mucho más intensos que en nuestro planeta.

En invierno, en el cinturón libre de nieve adyacente al borde del casquete polar marciano, se producen vientos suficientemente fuertes para dar lugar a tempestades de polvo. En esa región, tanto el *Mariner 9* como el Viking fotografiaron varias tempestades de polvo. Mucho más impresionantes son las tempestades de polvo globales, una de las cuales ocultó totalmente el planeta cuando el *Mariner 9* llegó y entró en órbita marciana en noviembre de 1971. Antes de las misiones Viking se creía que las tempestades de polvo globales se producían regularmente, una vez cada año marciano, hacia la época en que el planeta pasa a la mínima distancia del Sol. Muy recientemente, sin embargo, los vehículos Viking han detectado tempestades de polvo a escala casi global en una época del año marciano muy anterior. Es evidente que las grandes tempestades de polvo son más corrientes de lo que se había creído. Empezando por una tempestad en pequeña escala (o varias de ellas) en los trópicos meridionales, la tempestad global se extiende rápidamente, de tal modo que al cabo de unos días cubre la mayor parte del planeta.

¿Qué es lo que puede originar los fuertes vientos de tan enormes tempestades? Debe ser importante el hecho de que la órbita de Marte sea considerablemente más elíptica que la de la Tierra. La cantidad de energía solar que incide sobre Marte es alrededor de un 40 por ciento más grande cuando el planeta está en el punto más próximo al Sol que cuando está en el más alejado. La cifra correspondiente para la Tierra es solamente del 3 por ciento. Este hecho por sí sólo no puede, sin embargo, ex-

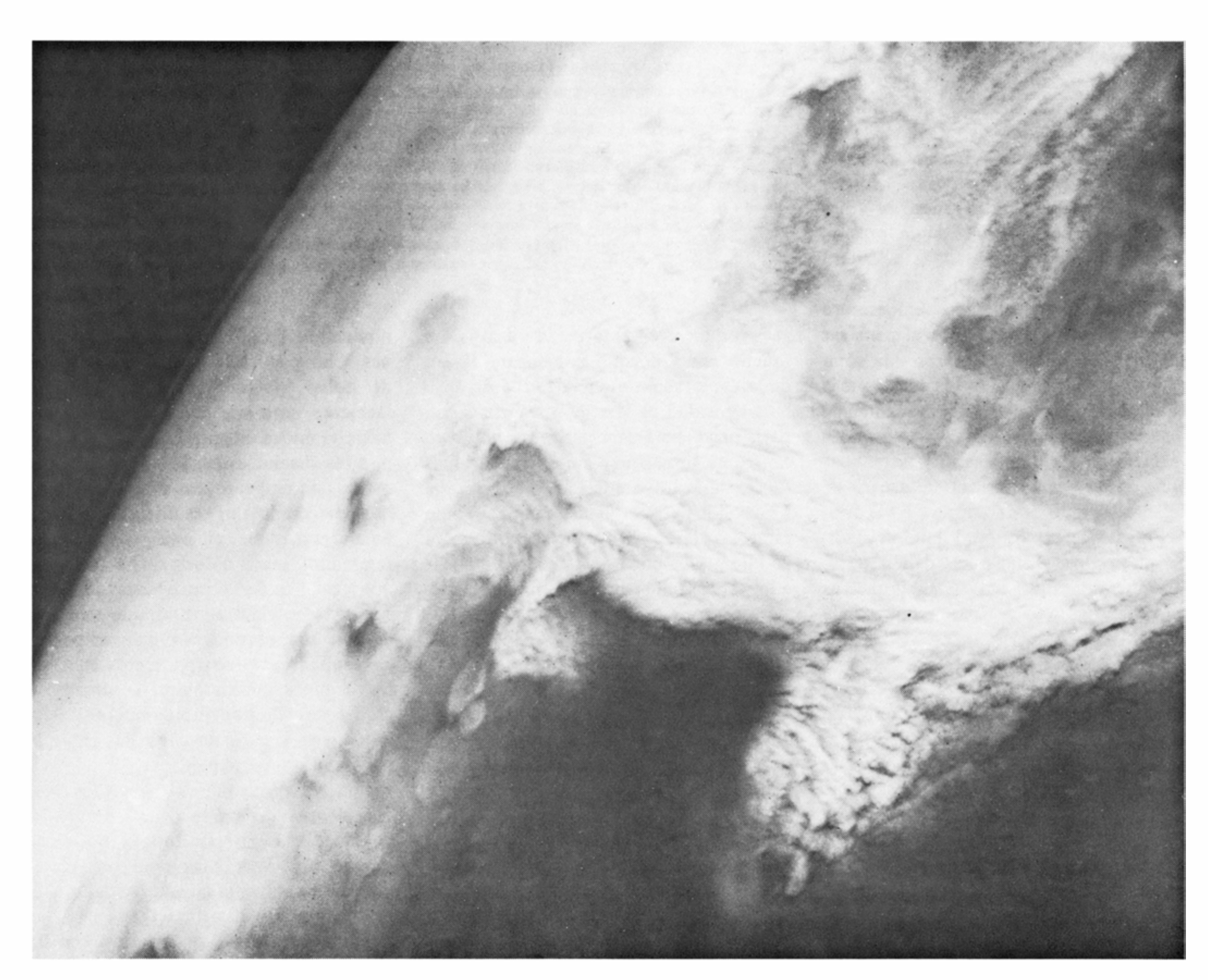
plicar las velocidades de viento que se requieren.

Se han propuesto cierto número de explicaciones, pero sólo voy a exponer aquella que encuentro más verosímil. La intensidad de los vientos globales de marea es muy sensible al calentamiento de la atmósfera. Ese calentamiento, a su vez, depende no sólo de la cantidad de energía solar que alcanza la superficie de Marte, sino también de la turbiedad de la atmósfera marciana: las partículas de polvo absorben luz solar y pueden calentar en forma efectiva y directa el gas circundante.

Si la atmósfera se encuentra suficientemente cargada de polvo, los vientos globales de marea pueden aumentar en intensidad hasta que sean casi todo lo fuertes que se requiere para levantar el polvo por sí mismos. Es de suponer que este proceso autorregenerador comience en regiones localizadas, donde los vientos de marea y los vientos topográficos locales se combinen para producir velocidades suficientemente altas como para incorporar a la atmósfera local una cantidad crítica de polvo. Conforme los vientos lo dispersan, aumenta el calentamiento de la atmósfera sobre amplias zonas, dando lugar a vientos de marea de la intensidad necesaria para levantar polvo sobre zonas todavía más vastas.

Parece claro que el polvo suspendido condiciona el comportamiento de la atmósfera en Marte más allá de cualquier cosa conocida en la Tierra. Incluso durante la estación en que los Viking se posaron recientemente sobre el suelo, estación en la cual la carga atmosférica de polvo era bastante baja, el cielo tenía un tinte rosáceo-amarillento en las fotografías tomadas por los vehículos, prueba de la existencia de una cantidad substancial de polvo dispersor de luz en la atmósfera. También parece necesario invocar el calentamiento solar de una atmósfera cargada de polvo para dar cuenta de variaciones diurnas en la presión a nivel del suelo tan grandes como las medidas en los lugares de toma. Parece, pues, que haya un eficaz sistema de realimentación positiva cuyas componentes sean los vientos, el polvo y el calentamiento atmosférico.

Sigue habiendo incertidumbres, pero el cuadro actual de los procesos que operan en la atmósfera de Marte presenta gran cantidad de detalle, a pesar de lo breve que es la historia de la exploración



VASTA TEMPESTAD DE POLVO MARCIANA fotografiada por el vehículo orbital *Viking 2* en febrero de este año desde un punto situado a unos 33.000 kilómetros por encima de la superficie del planeta y a 51 grados de latitud sur con una cámara que apuntaba al oeste-noroeste. El centro de la imagen está a 42,5 grados de latitud sur y 108 grados de longitud oeste, lo que coloca la tempestad en la región próxima a la Fosa Claritas y la Fosa Thaumasia, entre 20 y 45 grados de latitud sur. En unos pocos días, la tempestad se había extendido a la mayor parte del hemisferio sur, y duró varias semanas a partir de entonces. Cuando empezó la tempestad, el hemisferio sur marciano estaba a comienzos de primavera. Antes, se creía que tales tempestades de polvo se daban sólo en la época vecina al solsticio de verano del hemisferio sur.

del planeta desde cerca. Las fotografías y las mediciones de los Mariner y Viking han revelado muchas semejanzas sorprendentes entre la atmósfera marciana y la que conocemos en la Tierra. La química, los procesos dinámicos y la evolución atmosférica están estrechamente relacionados en ambos planetas. El esfuerzo que se hace para comprender esos procesos interrelacionados en otro planeta está ayudando a añadir perspectiva y amplitud a los estudios del nuestro.

¿Podemos decir algo a estas alturas sobre el clima marciano en el pasado? Y en concreto, los grandes canales y otras manifestaciones evidentes de erosión por agua, ¿indican acaso que Marte tuvo en alguna época un período dilatado de clima húmedo y cálido? Creo que hay pruebas suficientes para considerarlo improbable. Resulta que la cantidad total de gases escapados del interior ha sido en Marte muy inferior a la de la Tierra. Aun cuando la masa total surgida hubiese sido 100 veces la actual masa atmosférica o incluso superior, la disponibilidad de vapor de agua en la atmósfera puede que no haya sido muy superior a la de ahora. El factor que limita la cantidad de vapor de agua en la atmósfera marciana es la baja temperatura. Incluso una atmósfera de densidad próxima a la de la terrestre, pero compuesta principalmente de dióxido de carbono y nitrógeno, no pudo haber sido mucho más caliente que la actual y, por tanto, no pudo haber contenido mucho más vapor de agua.

Puede pensarse que el Sol mismo fuese más caliente cuando el agua erosionó las configuraciones que ahora vemos en Marte, pero parece más probable que sean el resultado de acontecimientos cataclísmicos tales como impactos de meteoritos o erupciones volcánicas. Es probable que la corteza marciana haya contenido grandes cantidades de agua, ya sea congelada o combinada químicamente, durante un largo tiempo, de modo que cualquier calentamiento brusco pudiese haber dado lugar a inundaciones locales.

Tal cataclismo, pues, podría haber producido también breves, e intensas, tormentas locales, en las que la súbita vaporización del agua fuese seguida rápidamente por lluvias torrenciales. Es posible que inundaciones cataclísmicas análogas se hayan producido en las primeras fases de la historia de la Tierra, antes de que hubiera océanos. Pero antes de que tengamos respuestas más precisas habrá que esperar a la próxima fase en la exploración de Marte.

Hepatitis vírica

En los últimos años se han producido importantes avances en la epidemiología e inmunología de esta enfermedad debilitante. Pronto se dispondrá de una vacuna que ofrezca protección contra una de sus formas, la hepatitis B

Joseph L. Melnick, Gordon R. Dreesman y F. Blaine Hollinger

La exigencia primordial a la hora de preparar una vacuna contra una enfermedad vírica fue, desde siempre, disponer de una gran cantidad del virus; en nuestro siglo, la tarea fundamental de los investigadores que han tratado de desarrollar una vacuna se ha centrado en poner a punto un buen sistema, tal como los embriones de pollo o los cultivos celulares, para la propagación del virus. Hay que destacar que, aún sin haberse obtenido todavía el crecimiento del virus en el laboratorio, las vacunas que se están ensayando en la actualidad podrán someter a nuestro control una enfermedad importante como es la hepatitis *B*. Las vacunas se basan en componentes de lo que parece ser una envoltura o capa externa del virus.

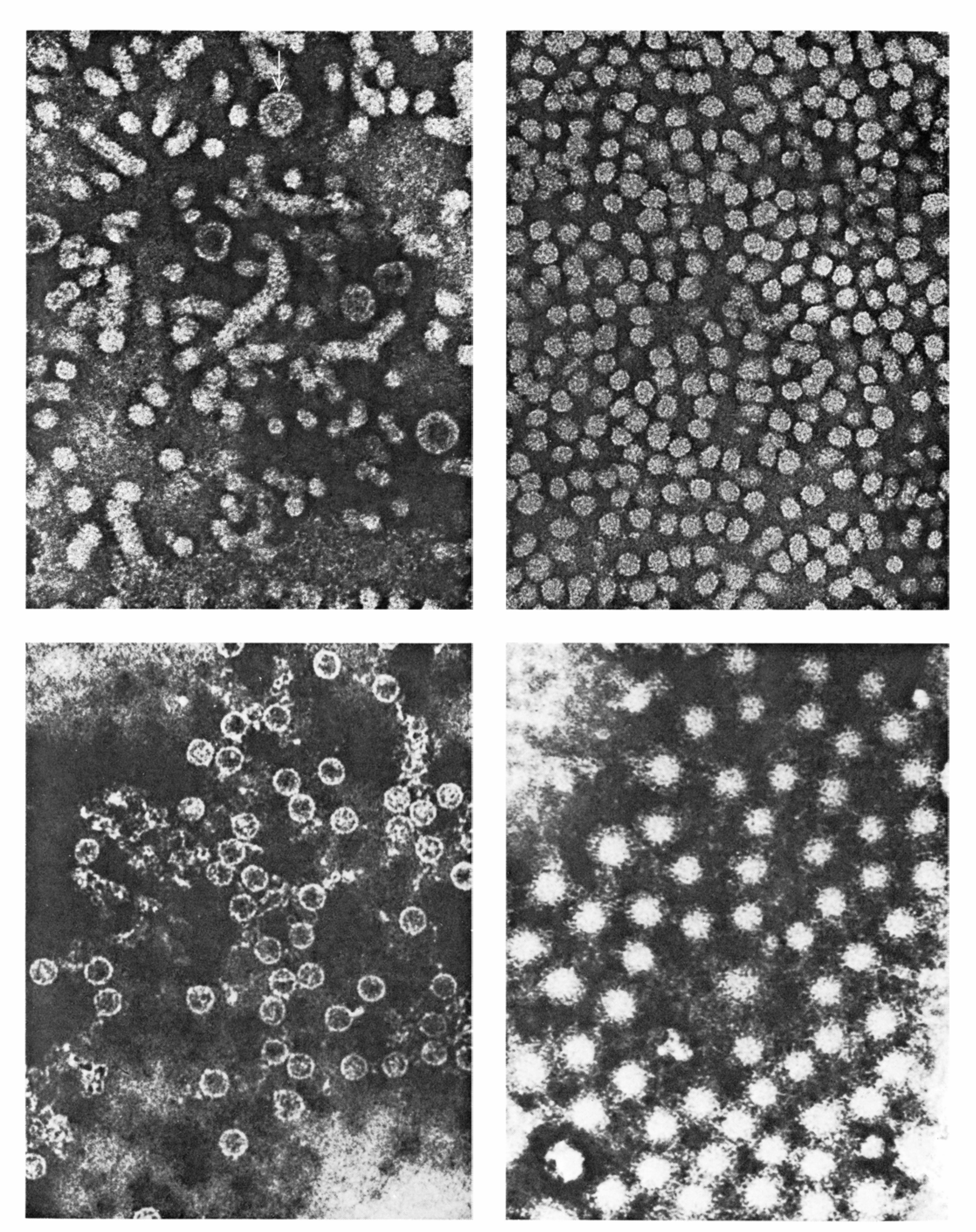
No se conoce ni la función precisa de estos componentes ni cómo ni cuándo se asocian al virus, pero se les encuentra en gran cantidad en la sangre de pacientes afectos de hepatitis *B* y en la de individuos que son portadores asintomáticos de la enfermedad. Por medio de la estimulación de la producción de anticuerpos altamente específicos, confieren inmunidad contra el virus como tal. En la revisión general que sigue, nos centraremos en el estudio de los caminos que nos han conducido hasta las puertas de la obtención de una vacuna contra la hepatitis *B*.

La expresión hepatitis vírica se emplea generalmente para referirse a infecciones hepáticas agudas causadas por uno de al menos tres diferentes tipos de virus. El signo más notable de estas enfermedades es la ictericia, coloración amarillenta de la piel y de la esclerótica debida al depósito de pigmentos biliares que el hígado ha sido incapaz de retirar de la circulación. Una forma importante es la hepatitis tipo *A*, o hepatitis de corta incubación, que antes se denominaba "hepatitis infecciosa" (y también, ictericia epidémica o ictericia catarral aguda). El virus responsable se transmite principalmente por la entrada de material de las heces de una persona infectada en el tracto digestivo de otra persona susceptible, pero en la actualidad se sabe que también puede transmitirse por otras vías. La otra forma importante de hepatitis vírica es la hepatitis *B*, o hepatitis de larga incubación, que se conocía anteriormente con los términos de hepatitis sérica (ictericia sérica homóloga, ictericia de transfusión o hepatitis de aguja). Como se deduce de esas denominaciones antiguas, la hepatitis *B* se transmite por la inoculación a través de la piel de sangre infectada o de componentes de ésta; recientes estudios, sin embargo, ponen de manifiesto que el virus puede transmitirse también por otras vías. Además de estas dos infecciones bien caracterizadas, se ha podido detectar una nueva forma de hepatitis asociada a las transfusiones sanguíneas, con sintomatología clínica y tiempo de incubación semejantes a los de la hepatitis *B*. Parece ser causada por un agente o agentes no relacionados con los virus de la hepatitis *A* o *B*; convencionalmente, nos referimos a ella como hepatitis *C*, aunque en su producción podrían hallarse involucrados varios agentes.

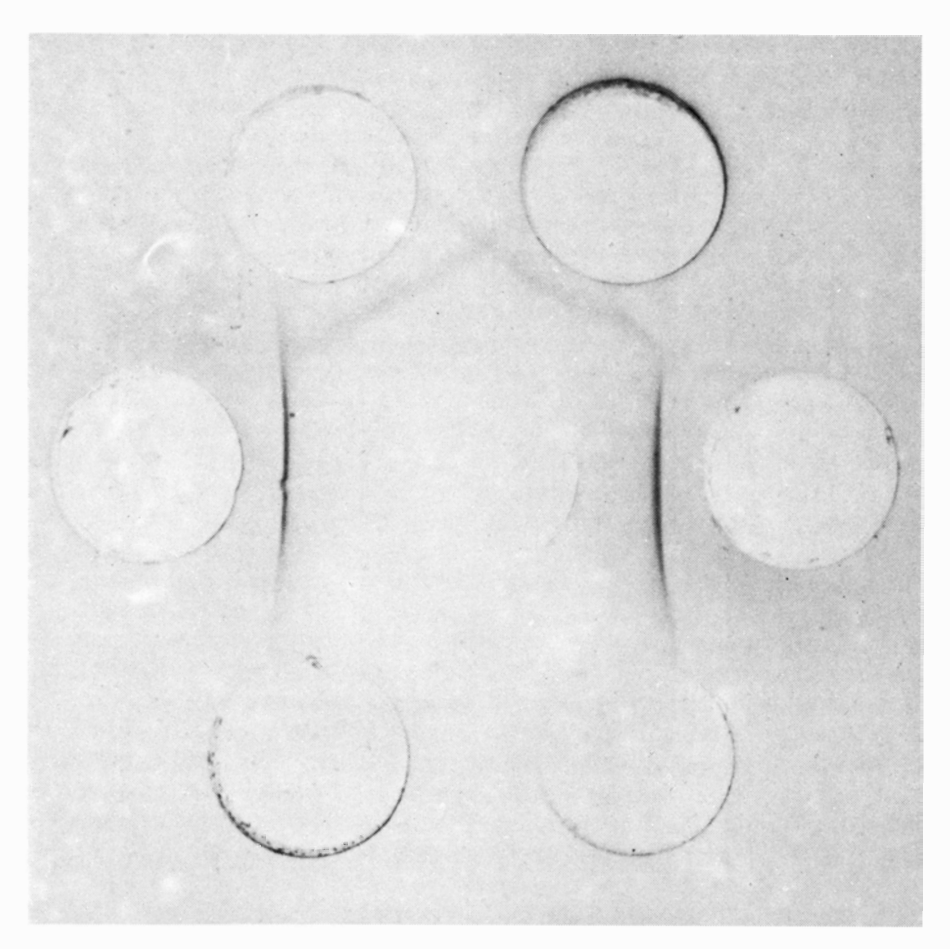
El año pasado se describieron en Estados Unidos cerca de 60.000 casos de hepatitis aguda; de ellos, un 60 por ciento fueron diagnosticados como hepatitis *A*, un 25 por ciento como hepatitis *B* y el 15 por ciento restante correspondió a un tipo de hepatitis sin especificar. Muchos casos son tan benignos que pasan inadvertidos; en otras ocasiones –y a pesar de detectar la enfermedad– el médico es incapaz de diagnosticarla; el número real de casos se ha estimado es unas diez veces mayor que el número descrito, lo que hace de la hepatitis una enfermedad altamente significativa en términos de incapacitación humana y costo público. En el caso de la hepatitis *B*, hay pacientes recuperados que se convierten en portadores del virus y los hay que desarrollan una forma crónica de la enfermedad. Estados Unidos cuenta hoy con un millón de portadores de la hepatitis *B*, aproximadamente.

Hasta los primeros años de la década de los sesenta no se había caracterizado a los dos virus responsables de las enfermedades que ahora conocemos como hepatitis *A* y hepatitis *B*; sólo se sabía que se trataba, al parecer, de dos formas distintas de enfermedad, una con un período de incubación mayor que la otra. Saul Krugman, de la New York University School of Medicine, y sus colaboradores habían acumulado y preservado suero de pacientes con cada una de las distintas formas de la enfermedad en cuestión.

Observaron que los niños expuestos al suero de pacientes de hepatitis de corta incubación desarrollaban regularmente una hepatitis clínica al cabo de unos treinta días, mientras que los niños a los que se les había administrado suero de pacientes con hepatitis de larga incubación desarrollaban sus primeros síntomas al cabo de unos sesenta días; cuando el suero de tales nuevos pacientes se administraba a otros niños, éstos –a su vez– desarrollaban la forma correspondiente de la enfermedad. La infección con un tipo de hepatitis confería inmunidad contra este mismo tipo, pero no contra el otro. Krugman y Joan P. Giles demostraron posteriormente que si se inactivaba, mediante calor, el suero de pacientes afectos de hepatitis *B*, y luego se inoculaba, protegía contra la



EL VIRUS DE LA HEPATITIS B y sus subunidades ampliadas unas 200.000 veces en tres microfotografías electrónicas hechas por Guy A. Cabral. En la sangre de los pacientes con hepatitis *B* pueden observarse habitualmente tres tipos de partículas *(arriba, izquierda)*: esferas de unos 22 nanómetros de diámetro, filamentos de 22 nanómetros de ancho y el virus de la hepatitis *B* *(flecha)*, conocido por partícula Dane. Las esferas de 22 nanómetros pueden ser purificadas *(arriba, derecha)*; constituyen la fuente principal del material para la preparación de la vacuna. Los hepatocitos de los pacientes con hepatitis *B* contienen partículas de 26 nanómetros *(abajo, izquierda)*, los núcleos no encapsulados del virus de doble pared. El virus de la hepatitis *A*, que no está relacionado con las partículas de la hepatitis *B*, está aumentado unas 200.000 veces *(abajo, derecha)* en ella se observan inmunoagregados en los que las partículas víricas están ligadas por anticuerpos específicos.



EL METODO DE GEL-DIFUSION, desarrollado por Orjan Ouchterlony, del Instituto Carolino de Estocolmo, detecta el antígeno de superficie de la hepatitis *B* en el suero sanguíneo y pone de manifiesto la existencia de subtipos diferentes de antígeno. Se labraron siete pocillos en una placa de agar sobre un vidrio (aumentado 7 veces aproximadamente). El suero que contiene el anticuerpo contra el antígeno de superficie se coloca en el pocillo del centro; el plasma perteneciente a seis individuos problema se coloca en los pocillos periféricos. Los antígenos en los sueros problema y el anticuerpo, en el centro, difunden radialmente a través del gel y, cuando se encuentran, se forman líneas de precipitación. Esta placa de Ouchterlony indica que cuatro de los sueros contienen el antígeno de superficie, tres pertenecientes a un subtipo y el cuarto suero a otro subtipo. Como puede apreciarse, el origen de la precipitación se explica en la ilustración de la página siguiente.

infección por virus de la hepatitis *B* a aquellos niños a quienes se les había administrado. Los experimentos demostraron que los agentes causales de las dos formas diferentes de hepatitis eran dos virus distintos, y apuntaban al mismo tiempo la posibilidad de obtener vacunas contra ellos a condición de que los virus se pudieran aislar o propagar.

La serie de acontecimientos que se sucedieron hasta llegar al descubrimiento del virus de la hepatitis *B* fue bastante complicada. Comenzó en 1963, cuando Baruch S. Blumberg, del Institute for Cancer Research, de Filadelfia, se encontraba examinando miles de muestras sanguíneas de diversas poblaciones en el marco de un plan de estudio de la variabilidad genética de las proteínas séricas, y no de la hepatitis. Observó que la muestra de un suero de un aborigen australiano reaccionaba con un anticuerpo presente en el suero de un paciente americano que padecía hemofilia. Este paciente, que había recibido múltiples transfusiones sanguíneas, había producido, posiblemente, anticuerpos como respuesta a una previa exposición al antígeno "Australia". ¿En qué consistía tal antígeno? Estudios subsiguientes demostraron que éste se mostraba con rareza en poblaciones de Estados Unidos o de Europa Occidental, siendo mucho más común en algunas poblaciones africanas y asiáticas. Con el tiempo, Blumberg y colaboradores dieron con pruebas que relacionaban el antígeno con lo que entonces se conocía por hepatitis sérica. En 1968, otros investigadores, especialmente Alfred M. Prince, del New York Blood Center, y K. Okochi, del Tokyo University Hospital, confirmaron que el antígeno Australia se hallaba de manera específica en el suero de los pacientes afectos de hepatitis *B*: el antígeno Australia era un antígeno de la hepatitis *B*.

Esta importante reacción antígeno-anticuerpo se detectó y estudió por medio de una técnica de gel-difusión. En esta técnica, se colocan dos sueros en pocillos labrados en un gel de agar, de suerte que se permita la difusión del uno hacia el otro: si uno de ellos contiene un antígeno, y el otro suero contiene un anticuerpo contra dicho antígeno, se forma una línea visible en el lugar donde antígeno y anticuerpo se encuentran y se combinan para formar un precipitado. Se observó que si el suero de dos pacientes reaccionaba con el anticuerpo de la hepatitis *B*, aparecía a veces un espolón secundario en la línea de precipitación, con lo que se ponía de manifiesto que había más de una forma de antígeno. Las partículas de la hepatitis *B* de cualquiera de estas dos personas contenían un antígeno "de grupo" común para todos los subtipos y al menos dos antígenos principales de subtipo.

Las partículas de la hepatitis *B* no tardaron en ser identificadas. Las microfotografías electrónicas revelaron que sueros que contenían el antígeno de la hepatitis *B* poseían variedad de partículas cuyas formas y tamaños eran similares a los de los virus animales conocidos: muchos pacientes presentaban hasta un billón de partículas por mililitro de sangre. La población más importante en la mayoría de muestras de sueros de hepatitis *B* estaba constituida por pequeñas partículas esféricas de unos 22 nanómetros (millonésimas de milímetro) de diámetro. También se encontraban formas largas, filamentosas, con idéntico diámetro y cuya longitud variaba entre algo menos de 100 y 700 nm. En 1970, D. S. Dane, del Middlesex Hospital Medical School de Londres, observó una nueva forma: se trataba de partículas semejantes a los virus con una doble capa: el diámetro externo media 42 nanómetros y 26 nanómetros el diámetro del núcleo central.

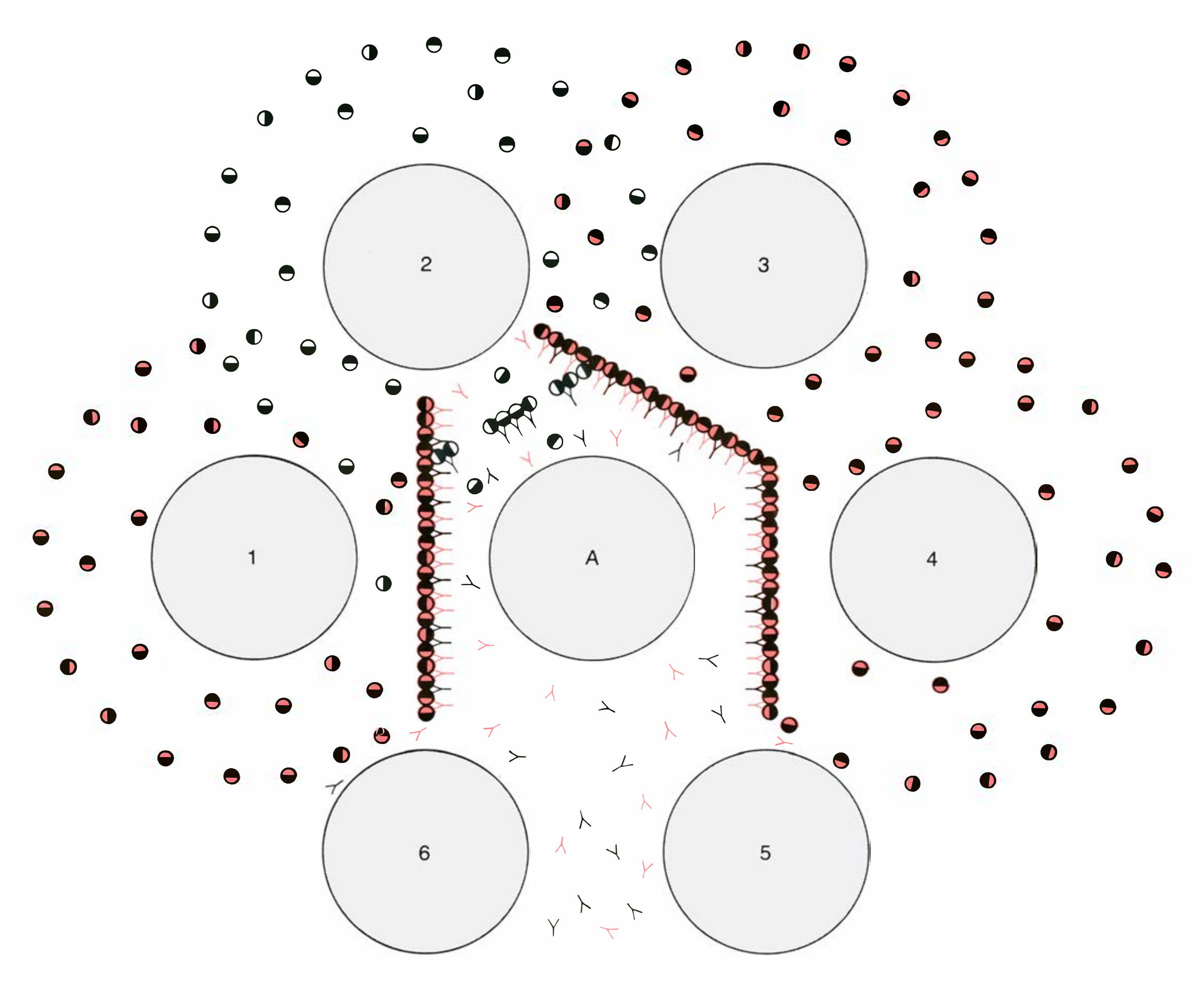
June D. Almeida, de los Wellcome Research Laboratories de Inglaterra, proporcionó la relación entre las anteriores observaciones de antígeno y anticuerpo y los datos electromicroscópicos: puso de relieve que, cuando un suero contenía anticuerpos contra hepatitis *B*, podía reaccionar con el suero que contenía las diferentes partículas antigénicas, produciendo la aglutinación de los tres tipos de partículas: esféricas, de 22 nanómetros, filamentosas y las partículas "Dane", que median 42 nanómetros.

Si se trataba a las muestras sanguíneas con un detergente para degradar la capa externa de la partícula Dane, el núcleo

(core) desnudo se mostraba incapaz de aglutinarse con las otras partículas. En otras palabras, la envoltura de las partículas Dane estaba antigénicamente relacionada con las partículas esféricas de 22 nanómetros y con las filamentosas, pero el núcleo de la partícula Dane era antigénicamente diferente de dichas formas. El antígeno asociado a las esferas, los filamentos y la envoltura se designó con el simbolismo HBsAg, o antígeno de superficie de la hepatitis *B*, y el antígeno del núcleo con el simbolismo variante HBcAg, o antígeno del núcleo de la hepatitis *B*.

Se han desarrollado métodos sensibles para la medición de cantidades muy pequeñas de estos dos antígenos: uno de los empleados con mayor amplitud es el radioinmunoensayo en fase sólida, en el que los complejos antígeno-anticuerpo pueden ser cuantificados en un contador de radiactividad, una vez marcados con isótopos radiactivos. Ese método es unas 4000 veces más sensible que la gel-difusión.

La siguiente etapa consistía en purificar y caracterizar las partículas completas de 22 nanómetros y ver si éstas tenían las características de un virus animal. Las partículas fueron purificadas en nuestro laboratorio en el Baylor College of Medicine y en otros laboratorios recurriendo a distintos procedimientos de centrifugación, siendo luego analizadas biofísica y bioquímicamente. Se halló que dichas partículas tenían una baja densidad, lo que significaba que contenían lípidos y proteínas. Análisis posteriores revelaron la presencia de hidratos de carbono, pero no la de ácidos nucleicos, material genético que debería estar presente en caso de tratarse de un virus. Nuevas mediciones, más precisas, mostraron que existía una amplia variación en su diámetro, entre 1 y 25 nanómetros. La composición aminoácida de la proteína parecía diferir considerablemente de la de los virus animales bien caracterizados. Además, las partículas tenían una naturaleza bioquí-



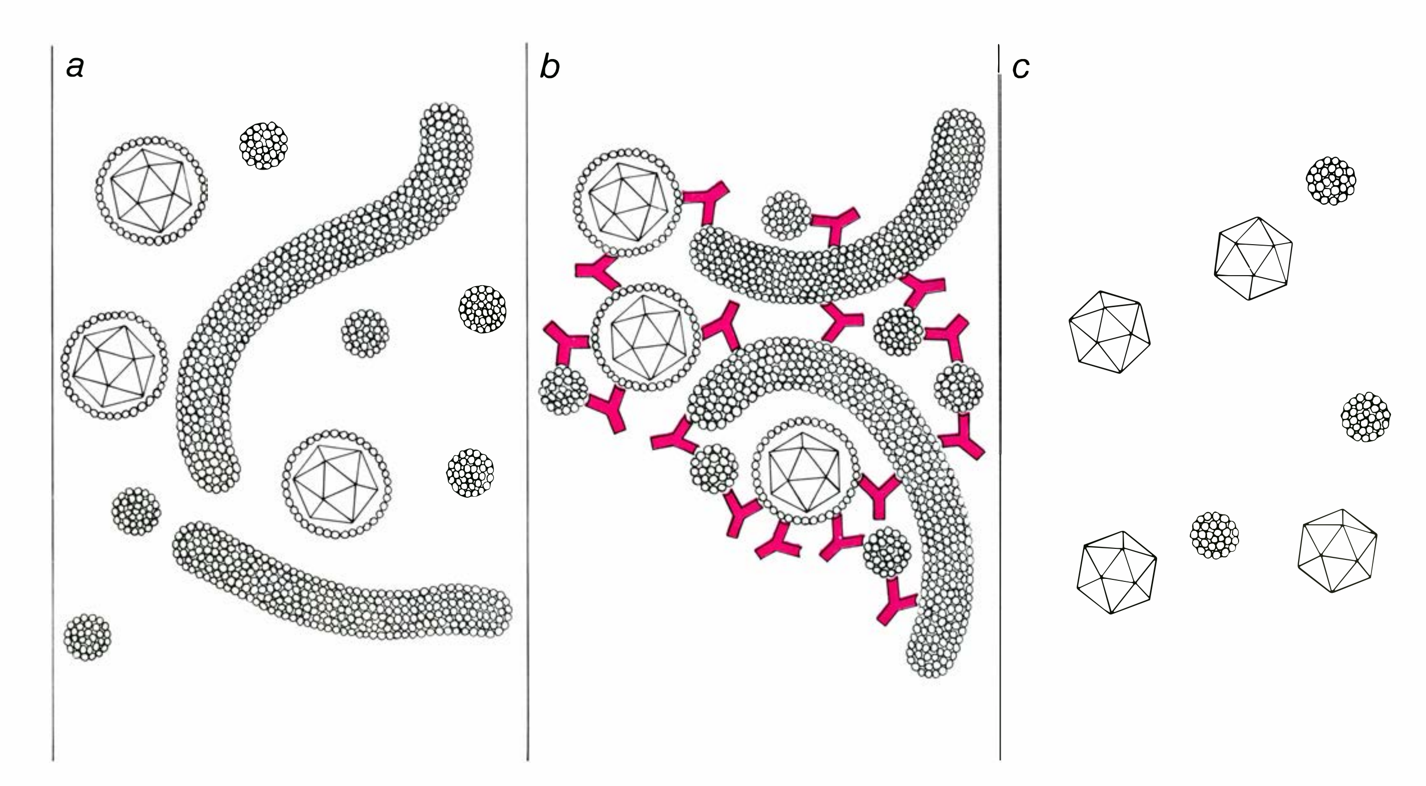
LAS LINEAS DE PRECIPITADO de la fotografía de la página anterior se forman por la reacción entre el antígeno *(pequeños círculos)* y el anticuerpo (en forma de *Y*). Los sueros de los pocillos *1, 2, 3* y *4* contienen antígeno de superficie; los pocillos *5* y *6* son negativos. Todas las partículas positivas para el antígeno tienen el mismo determinante antigénico de grupo *(negro)*, pero los de los pocillos *1, 3* y *4* tienen un determinante particular de tipo *(color)* y las del pocillo *2* tienen un determinante de tipo diferente *(blanco)*. El antisuero del pocillo *A* contiene anticuerpos contra el determinante de grupo y contra un determinante de tipo *(color)*, pero no contra el otro. La línea que aparece entre los pocillos *A* y *2* es más débil que las líneas entre el pocillo *A* y los *1, 3* y *4*, ya que, en ausencia del antígeno que pudiese emparejarse en el pocillo *2*, no se pueden formar tantos inmunocomplejos que precipiten. Se observan dos espolones entre los pocillos *1* y *2*, y entre los pocillos *2* y *3*. La razón de ello es que el anticuerpo específico de tipo *(color)* continúa difundiendo excéntricamente (en ausencia del antígeno correspondiente en el pocillo *2*) hasta que se une a los antígenos tipoespecíficos; estos, según queda explicado en el texto, provienen de los pocillos *1* y *3*.

mica compleja: constaban de unos ocho polipéptidos (cadenas cortas de aminoácidos) distintos o glucoproteínas (complejos proteína-azúcar). Todos estos hallazgos sugirieron que las partículas esféricas no representarían tal vez al virus, sino a algún producto específico de éste o bien algún material vírico que había sido modificado por las células del huésped en las que el virus había proliferado.

Poco tiempo después de que las diferentes partículas seudovíricas fueran observadas en la sangre de pacientes con hepatitis *B*, los investigadores hallaron material relacionado antigénicamente en los hepatocitos de hígados infectados. El citoplasma de las células infectadas contenía el antígeno de superficie, en tanto que el antígeno del núcleo se detectaba solamente en el núcleo de las células. Los antígenos fueron visualizados por primera vez, tras el marcaje del anticuerpo con un colorante fluorescente, bajo radiación ultravioleta. Su presencia se confirma con el microscopio electrónico; se observó que los hepatocitos infectados poseían en el núcleo el componente interno (core), de 26 nanómetros de diámetro, de las partículas Dane, y que dichas partículas no estaban presentes en los hepatocitos de pacientes afectos de otras hepatopatías.

William S. Robinson y Paul Kaplan, de la Stanford University School of Medicine, establecieron que las partículas nucleares aisladas de la sangre contenían un ADN de doble hélice y su propio ADN polimerasa (que es el enzima necesario para la replicación del ADN y, por tanto, para la multiplicación del virus). Nosotros purificamos las partículas nucleares a partir de hepatocitos infectados y hallamos que éstas también contenían ADN de doble hélice y el enzima polimerasa. El peso molecular y la densidad de las partículas nucleares eran los característicos de ciertos virus animales y se comportaban como partículas víricas cuando se someten a electroforesis. La mayoría de las preparaciones de partículas nucleares examinadas comprendían tres polipéptidos diferentes, lo que es típico de la cápside o cubierta proteica de muchos virus. Todos estos hallazgos juntos conducen a la conclusión de que la partícula Dane es un virus, con un núcleo interno y con una capa externa similar a la de los virus dotados de cápsula. El año pasado, la OMS (Organización Mundial de la Salud) recomendó que se considerase a la partícula Dane como el virus de la hepatitis *B*.

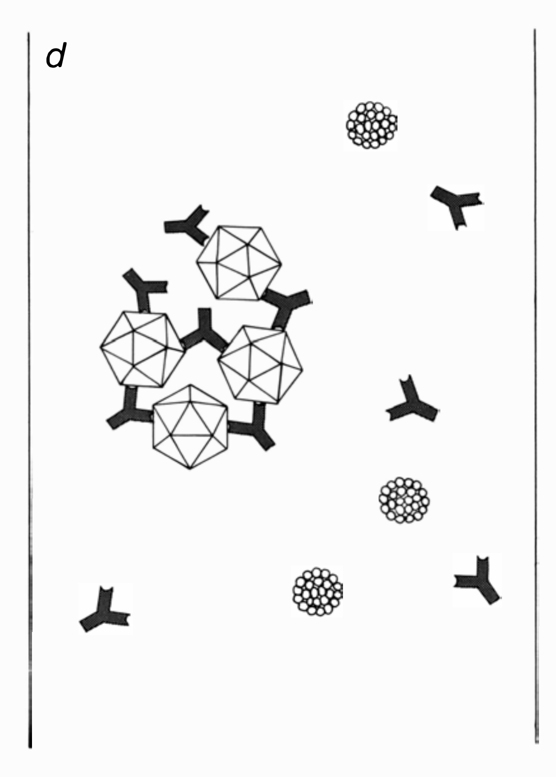
Además de los sistemas antígeno-anticuerpo de la superficie y del núcleo, se descubrió un tercer sistema en 1972, en el Laboratorio Bacteriológico Nacional de Estocolmo, cuyos componentes han recibido la denominación de antígeno *e* de la hepatitis *B* y anticuerpo anti-*e*. Al parecer el antígeno *e* está íntimamente relacionado con la partícula vírica encapsulada y con su capacidad de infección. Dicho antígeno se detecta solamente en individuos portadores del antígeno de superficie; su presencia está correlacionada con otros dos parámetros de infectividad: gran número de partículas víricas y actividad polimerasa del virus. Entre los portadores crónicos del antígeno de superficie de la hepatitis, aquellos que también son portadores del antígeno *e* desarrollan hepatopatía activa con una frecuencia mayor que aquellos otros cuyo suero presenta el anticuerpo anti-*e*. De la misma manera, la transmisión de la hepatitis a neonatos es más frecuente cuando la madre contiene en el suero el antígeno *e* que cuando contiene el anticuerpo anti-*e*; por último, la transfusión de sangre que contiene el antígeno *e* produce hepatitis *B* con mayor frecuencia que la



VARIAS PARTICULAS DE LA HEPATITIS B se relacionan y diferencian entre sí en cuanto a su antigenicidad. Las formas observadas en la sangre del paciente son las esferas de 22 nanómetros de diámetro, los filamentos de 12 nanómetros de ancho y el virus propiamente dicho (partículas Dane), que consta de una envoltura externa más un núcleo interno (*a*). Estas tres formas se caracterizan por la presencia en ellas del antígeno de superficie de la hepatitis *B*, como se indica por el hecho de que el anticuerpo (*color*) contra este antígeno se combine con todas ellas y las aglutine (*b*). La adición de un detergente a las partículas víricas elimina el antígeno de superficie de ellas, dando lugar

transfusión de sangre dotada del anticuerpo anti-*e*. En otras palabras, la presencia del antígeno *e* en la sangre de un portador del antígeno de superficie indica que dicha sangre es particularmente infecciosa; la presencia del anticuerpo anti-*e* significa que es mucho menos infecciosa. Sin embargo, sigue desconociéndose el locus específico del antígeno *e*.

A pesar de los datos que se van acumulando respecto al virus de la hepatitis *B* y a sus antígenos, todavía no se ha cultivado *in vitro* el virus como tal. Afortunadamente existe una posibilidad alternativa. Se demostró que los pacientes recuperados de una hepatitis *B* quedaban protegidos contra el antígeno de superficie por la presencia en su sangre de anticuerpos contra dicho antígeno. Los estudios prospectivos de hepatitis causada por transfusión de sangre que contenía el antígeno de superficie indicaban que ninguno de los receptores cuya sangre contenía el anticuerpo contra el antígeno de superficie desarrollaba una hepatitis *B* manifiesta, mientras que un número significativo de los que carecían del anticuerpo desarrollaban la enfermedad. Por otra parte, la presencia o ausencia de anticuerpos contra el antígeno de la partícula nuclear no parecía guardar relación alguna con la protección contra la infección vírica, propiamente dicha.



a los núcleos *(c)*. Estos núcleos en cuestión permanecen aglutinados por un anticuerpo que es diferente *(gris oscuro)*; a saber, el anticuerpo contra el antígeno del núcleo, que lo distingue de las partículas de superficie *(d)*.

De las investigaciones descritas se dedujeron dos conclusiones fundamentales. La primera, que el anticuerpo contra el antígeno de superficie protege contra la infección por el virus de la hepatitis *B*. La segunda, que las partículas esféricas de 22 nanómetros del antígeno de superficie podrían ser aisladas de la sangre de portadores aparentemente sanos en una forma altamente purificada y en cantidades lo suficientemente importantes como para proporcionar material para la preparación de una vacuna contra la hepatitis *B*. Un animal experimental para el ensayo de dicha vacuna debido a su susceptibilidad a la infección por el virus de la hepatitis *B* ha resultado ser el chimpancé. Animales inoculados con partículas purificadas de 22 nanómetros que habían sido tratadas con formaldehído desarrollaron anticuerpos contra el antígeno de superficie y posteriormente resistieron la infección cuando se les inyectó sangre conteniendo virus altamente infeccioso de la hepatitis *B*. Además, resultó que los animales se hacían inmunes contra el virus incluso cuando los anticuerpos iban dirigidos contra un subtipo de antígeno de superficie diferente de aquel con el que habían sido vacunados.

En el mes de diciembre pasado, Philippe Maupas, de la Universidad de Tours, presentó los primeros hallazgos de un estudio en el que voluntarios humanos habían sido vacunados con antígeno de superficie parcialmente purificado y tratado con formaldehído. Un 90 por ciento aproximadamente de los individuos vacunados desarrolló una respuesta inmune específica contra el antígeno. De 80 voluntarios vacunados que trabajaban en una unidad de hemodiálisis donde la incidencia de hepatitis *B* había sido persistentemente alta entre los pacientes y el personal, sólo un individuo desarrolló los síntomas de hepatitis *B*. Maupas y otros investigadores de ese país están llevando a cabo nuevas pruebas con las vacunas de antígeno de superficie.

No obstante, la vacunación con las partículas de 22 nanómetros debe efectuarse con precaución. En este sentido hay dos observaciones preocupantes: *a)* la complejidad de los componentes virales que, según parece, derivan de una infección por el virus de la hepatitis *B* es mayor de la que se podría esperar de acuerdo con la cantidad de información genética presente en el ADN de la partícula nuclear (core): *b)* la inoculación en animales pequeños con las partículas de antígeno de superficie produce una respuesta inmunológica vigorosa, mientras que los humanos (el huésped infectado de forma natural) y los chimpancés producen bajos niveles de anticuerpos contra el antígeno de superficie, incluso después de la infección. Estas observaciones sugieren que las partículas de 22 nanómetros podrían contener componentes celulares antigénicos (quizá ciertas proteínas hepáticas) derivados del huésped, y ello con preferencia a proteínas extrañas puramente asociadas con el virus. En virtud de lo cual, algunos individuos podrían desarrollar una respuesta autoinmune, a saber, contra sus propias proteínas hepáticas. Corroborando esta posibilidad, varios laboratorios han dado a conocer el siguiente hallazgo: pacientes con hepatitis *B* crónica activa desarrollan una inmunorrespuesta contra lipoproteínas (complejos de proteínas y lípidos) normales del tejido hepático.

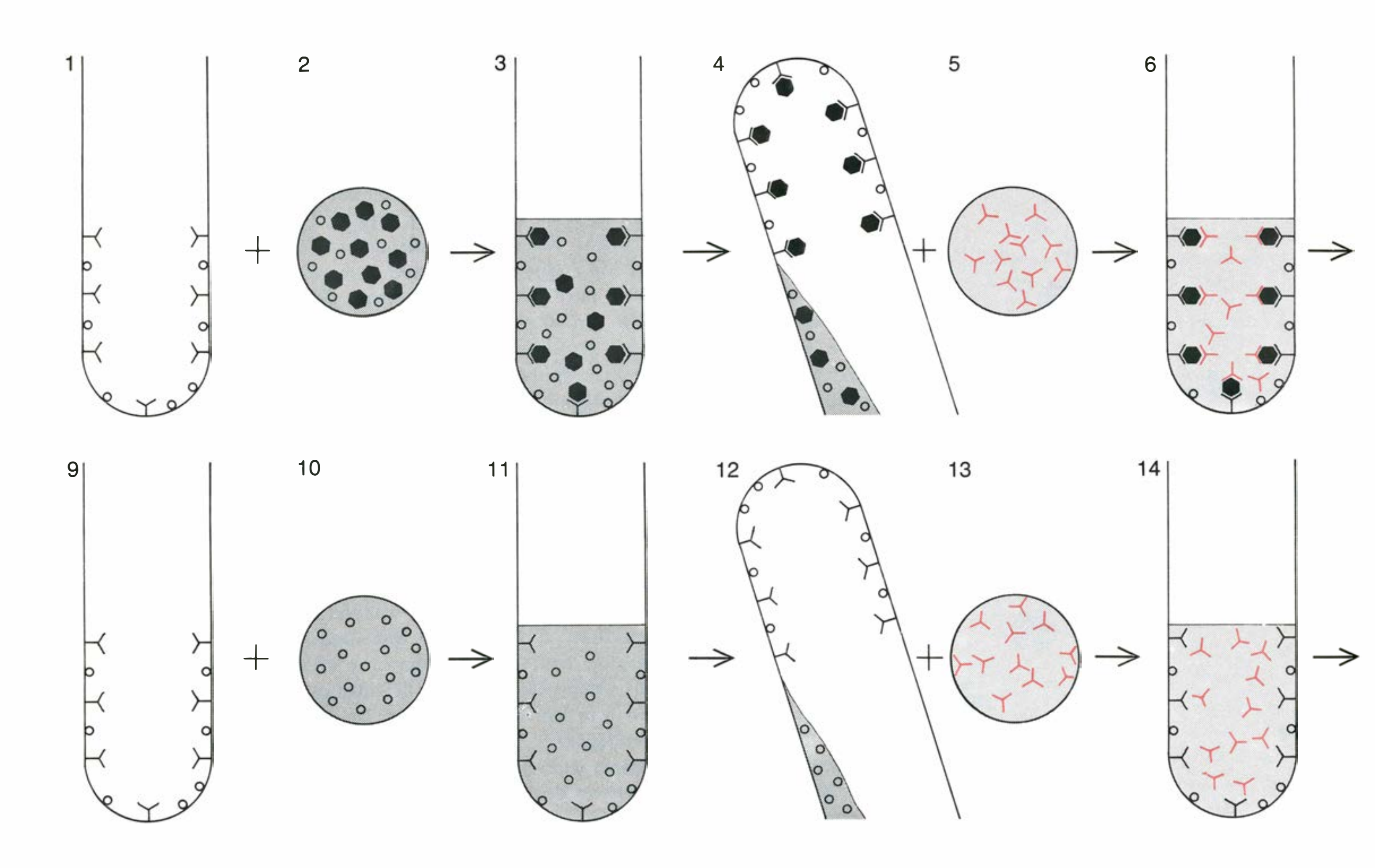
A la vista de dichos problemas, parece recomendable considerar el estudio de una forma más segura de vacunación. En este sentido, nuestro laboratorio y otros están investigando las propiedades inmunogénicas de los componentes polipeptídicos de las partículas purificadas de 22 nanómetros. Al parecer, la mayoría de los polipéptidos inducen una inmunorrespuesta con anticuerpos que reaccionan con la partícula de 22 nanómetros intacta. Se está experimentando cuidadosamente en animales para ver si desarrollan respuesta autoinmune contra componentes del hepatocito normal; sólo los polipéptidos que fracasaran en la inducción de dichas respuestas deberían ser considerados como candidatos para una vacuna. Las preparaciones de pequeños polipéptidos purificados tendrían la ventaja adicional de hallarse libres de genes celulares o virales que pudieran tener cierta capacidad infectiva. Si resulta que una vacuna formada mediante un solo polipéptido posee ventajas sobre otra preparada a partir de las partículas completas de 22 nanómetros, habrá que desarrollar una metodología económica y eficaz que permita preparar suficientes cantidades de polipéptido a partir de las partículas.

El virus asociado a la hepatitis *A* es totalmente distinto del virus de la hepatitis *B*. Nuestros conocimientos sobre él son aún muy escasos. S. M. Feintone y sus colegas, de los National Institutes of Health, fueron quienes primero lo lograron visualizar en heces obtenidas en la fase aguda de la enfermedad, padecida por voluntarios que habían sido infectados con suero de pacientes afec-

tos de hepatitis *A*. La partícula vírica mide 27 nanómetros de diámetro; a diferencia de lo que ocurre con el virus de la hepatitis *B*, no presenta cápsula externa. Se ha descrito un amplio rango de densidades para estas partículas víricas, lo que probablemente se debe a que las partículas diferentes contienen distintas cantidades de ácidos nucleicos; algunas de las cuales, en su pobreza, llegan a carecer absolutamente de ácido nucleico. Todavía no ha sido posible establecer cuál sea el material genético de estos virus. Más que ADN, como en el virus de la hepatitis *B*, algunos hallazgos preliminares sugieren que el virus de la hepatitis *A* es un ARN vírico, como en poliovirus y otros virus afines. Sin embargo, si nos atenemos a otras propiedades, como la estabilidad frente al calor, y a su amplio rango de densidades, había que inclinarse porque el virus se halle estrechamente relacionado con un grupo particular de pequeños virus que contienen ADN. Los estudios definitivos requerirán la obtención de una fuente importante del virus de la hepatitis *A* para su posterior análisis bioquímico: es decir, habrá que esperarse al desarrollo de una metodología que permita que el virus pueda crecer en el laboratorio.

La epidemiología de la hepatitis *A* y *B*, y los cambios serológicos observados en las dos enfermedades, difieren completamente, tal y como corresponde a dos infecciones causadas por dos virus distintos. No obstante, resulta extremadamente difícil diferenciar entre ambas enfermedades en los primeros estadios, a juzgar por los rasgos clínicos. En cualquier caso, el paciente típico desarrolla una sintomatología peculiar (malestar, dejadez, mialgias, anorexia, náuseas y dolores abdominales) y presenta febrícula. La orina deviene de color marrón y la piel adquiere una tez amarillenta: algunas veces, la sintomatología disminuye después de la aparición de la ictericia. (En algunos casos de hepatitis *B* hay artritis, vasculitis y afectación glomerular, a veces asociada a una erupción cutánea y sin ictericia; sin embargo, el diagnóstico de hepatitis *B* puede ser confirmado por el examen de la sangre para los diferentes marcadores serológicos y para la prueba bioquímica de la disfunción hepática, observándose un nivel notablemente elevado de los enzimas hepáticos conocidos como transaminasas.)

La recuperación suele ser completa en un plazo de dos a cuatro semanas para la hepatitis *A* y, de cuatro a ocho semanas, para la hepatitis *B*. Entre un 10 y un 15 por ciento de los pacientes con hepatitis *B* desarrollan una forma severa de la enfermedad. En algo menos del 2 por ciento de los casos, la enfermedad progresa hacia una insuficiencia importante y rápida del funcionalismo hepático: en estos casos, la mortalidad alcanza entre un 70 y un 100 por ciento de los pacientes. Para la hepatitis *A*, el grado de fatalidad es del 0,5 por ciento en todos los casos descritos; para la hepatitis *B*, asociada a transfusión, es del 3 al 4 por ciento y, para la hepatitis *B* por abuso de drogas, es del 0,8 por ciento. Obviamente estas cifras pueden resultar extremadamente elevadas, dada



EL ANTIGENO puede detectarse por la técnica de radioinmunoensayo en fase sólida. El anticuerpo específico *(Y, de color negro)* **contra el antígeno que se desea estudiar se adsorbe a un soporte sólido, que en este caso es la pared de un tubo de ensayo de plástico** *(1)*, **junto a otras proteínas no específicas** *(círculos abiertos)*. **Se añade la muestra** *(2)*; **si ésta contiene el antígeno** *(hexágonos negros)*, **el antígeno se combina con el anticuerpo** *(3)*. **El antígeno que no se ha unido se lava** *(4)*. **Para poder detectar el antígeno ligado, se añade nuevamente anticuerpo, pero en esta ocasión marcado con un isótopo radiactivo** *(Y, en color) (5)*. **Este se une al antígeno ligado, que entonces queda emparedado entre los dos an-**

la alta frecuencia de casos benignos asintomáticos que no se describen.

El tratamiento de la hepatitis viral es fundamentalmente sintomático, lo que quiere decir que la labor por hacer en el campo terapéutico es todavía muy importante. Los pacientes deben permanecer en cama y abstenerse de la ingestión de alcohol y de otros tóxicos hepáticos. Se han utilizado esteroides adrenocorticales, pero no se han mostrado eficaces en los casos no complicados, y si el cese de su administración es brusco pueden dar lugar a subsiguientes recaídas. En los casos más severos se han administrado esteroides y gammaglobulina sérica inmune contra la hepatitis *B*, y de nuevo se ha fracasado en la demostración de su eficacia terapéutica.

Hoy se está investigando en la terapia con interferón, que es la sustancia antiviral secretada por las células que sufren el ataque vírico. En las primeras pruebas en la Stanford University School of Medicine se ha registrado un modesto éxito en un limitado número de pacientes con hepatitis crónica *B*. Hubo supresión pasajera de los diferentes marcadores víricos presentes en la sangre, pero no se observaron efectos definitivos sobre la hepatopatía como tal. Los resultados parecen indicar que el interferón puede ser eficaz para limitar la infectividad de los individuos que son portadores del virus, e incluso en la irradiación de las infecciones crónicas que afectan a algunos portadores. La obtención de resultados definitivos requiere todavía el llevar a cabo más estudios del tipo del doble ciego.

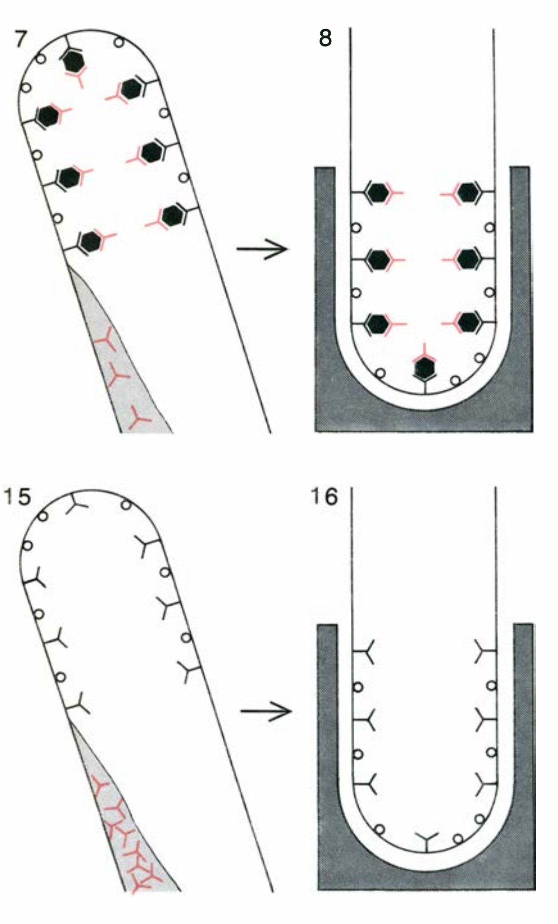
Los acontecimientos inmunológicos más importantes asociados con la hepatitis *A* son distintos de los de la hepatitis *B*. En la hepatitis *A*, el virus se multiplica probablemente a nivel intestinal, y luego se extiende al hígado por diseminación hematógena o linfática. Se ha llegado a descubrir la excreción fecal del virus o de los productos víricos, incluso dos o tres semanas antes de la instauración de la ictericia: según parece, la enfermedad encierra su máxima contagiosidad cerca del final del período de incubación (antes de la aparición de la sintomatología), es decir, cuando el virus adquiere su máxima concentración en las heces: tres o más semanas después de la instauración de los síntomas, las heces dejan de ser infecciosas. En el estadio precoz de la enfermedad, hay bajos niveles víricos en la sangre y en la orina, pero la enfermedad raramente se transmite por estas fuentes: tampoco se cree que la vía de contagio sea el estornudo. Varios estudios han mostrado una relación directa entre la prevalencia de la hepatitis *A* (valorada por la prevalencia del anticuerpo contra el virus) y el aumento de la edad o la higiene (personal o ambiental) deficiente, una correlación que es común a muchas enfermedades víricas que se producen por multiplicación en el tracto intestinal.

Los acontecimientos clínicos e inmunológicos que siguen a la infección por el virus de la hepatitis *B* reflejan una compleja interacción entre el huésped, el virus y algunos antígenos virales en particular. En un caso agudo típico, el antígeno de superficie aparece en la sangre uno o dos meses después de la infección: puede persistir durante unos días o incluso varios meses. Las partículas víricas, el antígeno *e* y la actividad polimerasa aparecen en la sangre junto con el antígeno de superficie. A medida que la infección progresa, los hepatocitos sufren necrosis y liberan a la sangre las transaminasas hepáticas: el hallazgo de gran concentración de transaminasas prenuncia los síntomas de la hepatitis. Aunque el antígeno de superficie es habitualmente la prueba de más valor en la hepatitis *B*, éste puede desaparecer de la sangre muy tempranamente: en dichos casos, el diagnóstico puede establecerse por la presencia de anticuerpos contra el antígeno del núcleo, que se hace detectable al mismo tiempo que los signos clínicos y bioquímicos. El anticuerpo contra el antígeno de superficie también se hace detectable cuando desaparece el antígeno, pero sólo si se utilizan técnicas muy sensibles.

En pacientes que desarrollan una hepatitis crónica o que se hacen portadores asintomáticos, el antígeno de superficie persiste en la sangre, su anticuerpo no aparece y se pueden detectar altas concentraciones de anticuerpo y antígeno del núcleo. En los pacientes con una hepatitis *B* crónica, la sangre contiene frecuentemente el antígeno *e*, la ADN polimerasa y partículas víricas, mientras que la sangre de portadores en apariencia sanos presenta con mayor frecuencia los anticuerpos anti-*e*. Si un individuo inmune contra la hepatitis *B* queda reexpuesto al virus, hay una rápida respuesta inmunológica con memoria, con aparición del anticuerpo contra el antígeno de superficie dentro de dos semanas: el antígeno de superficie como tal no se detecta y la persona no padece infección.

La incidencia de la hepatitis *B* en la población general es mayor en el grupo de edad comprendido entre los 15 y los 29 años, presumiblemente como resultado del abuso de drogas. La incidencia es mayor, sin embargo, en grupos específicos tales como los sometidos a hemodiálisis y quienes los atienden a éstos. Hasta hace poco, era muy frecuente en receptores de sangre o de plasma, pero la incidencia de la infección por el virus de la hepatitis *B* asociada a transfusiones ha disminuido casi en un 75 por ciento, ya que el estudio de los donantes, por medio de métodos sensibles, como el radioinmunoensayo, se ha hecho imprescindible.

La extensión natural de la hepatitis *B* por contacto, más que por punción cutánea, parece ser más común de lo que se venía creyendo, en particular en el caso de los compañeros sexuales de los portadores del virus (tengan éstos sintomatología crónica o no) y entre los residentes de instituciones muy nutridas. El hecho de que el antígeno de superficie se detecte regularmente en varias excreciones y secreciones corporales,

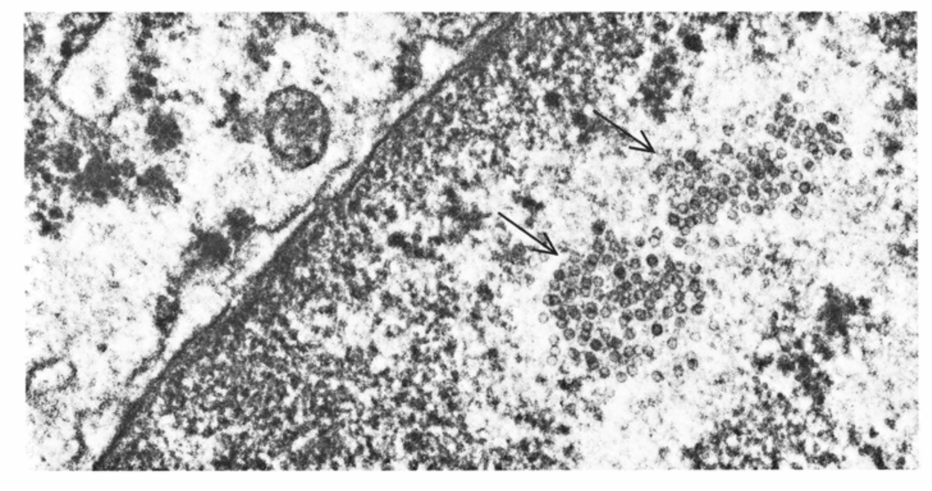


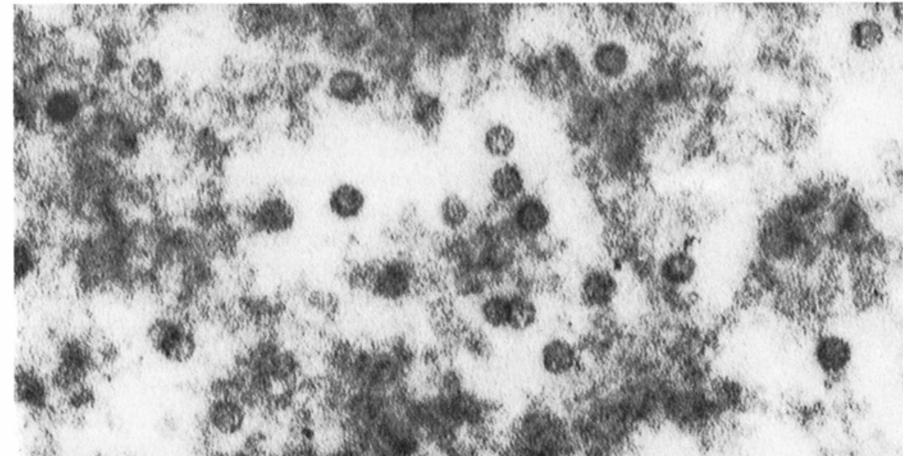
ticuerpos *(6)*. **El tubo se lava de nuevo, quedando sólo los emparedados de anticuerpo-antígeno-anticuerpo** *(7)*. **La radiactividad del tubo se mide en un contador** *(8)* **y se compara con la de una muestra que hace de control negativo** *(9-16)*.

apoya la idea de que la hepatitis *B* puede transmitirse por rutas que no conllevan la penetración a través de la piel. Dentro de una misma familia, un portador del virus puede transmitir la enfermedad fácilmente a otro miembro que utilice su toalla, navaja de afeitar o cepillo de dientes. En general, la hepatitis *B* se contagia con mayor probabilidad desde un paciente en la fase aguda de la enfermedad que desde un portador asintomático; la transmisión de la enfermedad desde un portador asintomático puede requerir la exposición repetida al agente infeccioso.

No todos los individuos que resultan expuestos padecen posteriormente la infección. De los que lo hacen, sólo una pequeña proporción desarrolla enfermedad clínicamente observable; su porcentaje depende ante todo de la dosis y de la vía de infección: pequeñas dosis del virus y la transmisión por una vía distinta de la cutánea resultan en un mayor período de incubación y en una menor probabilidad de sintomatología clara, y la ictericia sólo aparecerá, quizás, en un uno por ciento de los casos. Dosis mayores transmitidas a través de la piel pueden disminuir el período de incubación y aumentar la relación ictericia:no-ictericia en una proporción de uno es a cinco. Los que resultan infectados pero no desarrollan sintomatología, y los que se hacen portadores (definidos como aquellos cuyo nivel sanguíneo de antígeno de superficie permanece en los mismos niveles durante más de cuatro meses), constituyen una fuente significativa de contagio.

En varios estudios se ha observado un alto predominio del antígeno de superficie en individuos que tienen un carcinoma primario hepático: en un 80 por ciento de dichos pacientes en Taiwan (comparado con un 15 por ciento en la muestra control) y en un 35 por ciento de dichos pacientes en España (comparado con un uno por ciento en los controles). Similares cifras de antígeno de superficie se han registrado en varias zonas de los Estados Unidos en pacientes con carcinoma hepatocelular, desarrollado a partir de una cirrosis hepática no alcohólica. Se han avanzado varias hipótesis para explicar la asociación entre





NUCLEO DE LOS HEPATOCITOS de un paciente con hepatitis *B* crónica, que murió por enfermedad renal. Contiene partículas nucleares de la hepatitis *B*. Los grupos de partículas *(flechas)* están ampliados unas 55.000 veces *(arriba)*; las partículas de otro núcleo están aumentadas unas 200.000 veces *(abajo)*. Micrografía de Ferenc Gyorkey, del Veterans Administration Hospital, Boston.

VIRUS EN SANGRE
VIRUS EN HECES
TRANSAMINASAS
SINTOMATOLOGIA
CONCENTRACION RELATIVA DEL ANTICUERPO
INMUNOGLOBULINA *G* (IgG)
INMUNOGLOBULINA *M* (IgM)
NIVEL DE DETECCION
0 2 4 6 8 10 12
SEMANAS DESPUES DE LA EXPOSICION

LA HEPATITIS A tiene un período de incubación de cuatro semanas entre la exposición y la aparición de los síntomas, durante los cuales el virus y las transaminasas hepáticas se detectan en la sangre. Las curvas trazan la concentración del anticuerpo antivírico en dos tipos de inmunoglobulinas: la inmunoglobulina *M*, que es un anticuerpo de la fase aguda, y la inmunoglobulina *G*, que domina durante la convalecencia.

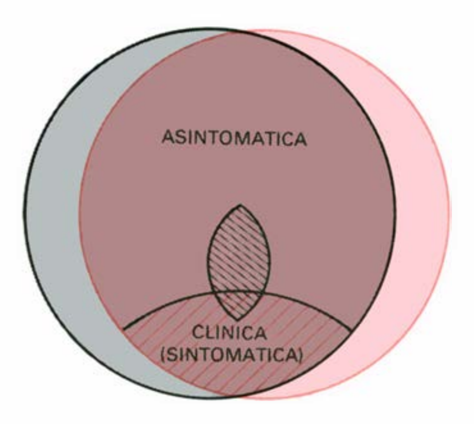
TRANSAMINASAS
SINTOMATOLOGIA
PERIODO DE INCUBACION
FASE AGUDA
CONVALECENCIA
PRECOZ
TARDIA
PRUEBAS DIAGNOSTICAS
ANTICUERPO CONTRA EL ANTIGENO NUCLEAR
ANTIGENO DE SUPERFICIE
ANTICUERPO CONTRA EL ANTIGENO DE SUPERFICIE
CONCENTRACIONES RELATIVAS DE LOS MARCADORES DE LA HEPATITIS *B*
ANTICUERPO CONTRA EL ANTIGENO NUCLEAR
ANTIGENO DE SUPERFICIE
VIRUS, ADN POLIMERASA, ANTIGENO NUCLEAR Y ANTIGENO *e*
ANTICUERPO CONTRA EL ANTIGENO DE SUPERFICIE
NIVEL DE DETECCION
0 1 2 3 4 5 6 7 8 9
MESES DESPUES DE LA EXPOSICION

LA HEPATITIS B tiene un período de incubación más largo que el de la hepatitis *A*. Los diferentes antígenos de la hepatitis *B* y sus anticuerpos se dejan observar y varían su concentración a lo largo de la infección; la eficacia de los procedimientos diagnósticos serológicos varía asimismo.

la hepatitis *B* y el cáncer hepático. Una de ellas sostiene que los pacientes con cáncer hepático o con cirrosis son particularmente susceptibles de la infección por el virus de la hepatitis *B*, o para comportarse como portadores crónicos. Otra, que la infección conduce a cirrosis hepática y ésta es el precursor real del cáncer. Por último, una tercera defiende que el virus de la hepatitis *B* puede transformar los hepatocitos que invade, es decir, que se trata de un virus oncógeno (influido, quizá, por factores genéticos, ambientales u otros). Si cualquiera de las dos últimas interpretaciones resulta ser la correcta, el control de la hepatitis *B* por inmunización debería reducir la incidencia de cáncer hepático.

Algunos pacientes transfundidos continúan desarrollando hepatitis viral a pesar del cuidadoso estudio de los donantes y de sus productos sanguíneos para la presencia de marcadores de la hepatitis *A* y *B* (sin que se manifestaran, obviamente, estos marcadores, en la sangre de los pacientes). En ausencia de métodos de identificación de los agentes productores de la hepatitis no-*A* y no-*B* (que ahora dan cuenta de más del 80 por ciento de las hepatitis post-transfusionales que se describen en los Estados Unidos), no podemos conocer mucho acerca de sus características inmunológicas ni de sus vías de transmisión. Sin embargo, ya podemos ir tomando nota de algunos puntos. Los patrones clínicos y epidemiológicos son similares a los de la hepatitis *B*, la enfermedad parece ser común y presentar también un estado de portador, y los receptores de sangre de donantes con un nivel de transaminasas elevado tienen presumiblemente –al menos– un riesgo aumentado de desarrollo de esta forma recientemente reconocida de hepatitis.

Hasta que no se disponga de nuevos resultados sobre las vacunas de la hepatitis *B*, el esfuerzo destinado a controlar la hepatitis viral dependerá de los métodos tradicionales, lo que quiere decir una buena higiene personal, medidas importantes de salud pública y el uso de la gammaglobulina específica. El lavado de las manos antes de las comidas o de la manipulación de los alimentos es probablemente el único procedimiento preventivo en el caso de la hepatitis *A*; por lo que se refiere a la hepatitis *B*, debe evitarse la exposición a cualquier objeto que pueda estar contaminado con sangre, desde una toalla prestada o una maquinilla de afeitar hasta una aguja hipodérmica. La sustitución de los bancos de sangre comerciales y de los donantes retribuidos por un sistema basado en la donación voluntaria exclusivamente reduciría la incidencia de los tipos *B* y *C* de hepatitis. El uso de la inmunoglobulina (lo que habitualmente conocemos como gammaglobulina) debe aplicarse a individuos cuyo contacto se produce en el seno de una familia o de una institución, pero no probablemente en los contactos rutinarios en el trabajo o en la escuela: la administración de la gammaglobulina seis semanas después de la exposición o del desarrollo de la sintomatología no sirve para nada. En el caso de la hepatitis *B*, la gammaglobulina standard debe reservarse para los casos en que se haya inoculado accidentalmente una pequeña cantidad del virus. Recientemente se ha producido una gammaglobulina altamente concentrada contra la hepatitis *B*, cuyos resultados preliminares han sido prometedores en algunos grupos con alto riesgo de infección.



LA INFECCION POR LA HEPATITIS B se representa aquí como la intersección del virus *(color)* **con la población huésped** *(gris)*. **La mayoría de las infecciones son inaparentes, pero algunas son francamente clínicas** *(rayado en color)*. **En cualquier caso, algunos individuos infectados devienen portadores** *(rayado en negro)*.

Nuestros conocimientos actuales sobre la naturaleza y manipulación de los antígenos de la hepatitis *B* permiten abrigar la esperanza de que los ensayos de la vacuna que se están realizando darán resultados positivos y que, dentro de los próximos dos años, dispondremos de una vacuna para la inmunización de los individuos que se hallan expuestos con frecuencia a la hepatitis *B*. Los avances de la última década en los conocimientos relacionados con la hepatitis *B* han sido realmente muy notables, y las técnicas elaboradas para la investigación de la hepatitis han proporcionado nuevos instrumentos que sirven en el estudio de un amplio espectro de enfermedades humanas que se sospecha son de origen vírico y para los que las técnicas de trabajo habituales resultaron ineficaces.

Ciencia y sociedad

Leptones ambivalentes

Desde hace largo tiempo, el electrón y el muón constituían una extraña pareja. Son tan parecidos –se diferencian sólo en su masa– que es difícil entender por qué razón habrían de existir como dos partículas independientes. Podría bastar con cualquiera de ambas partículas.

Por otra parte, puesto que hay dos partículas, no hay motivo aparente para que no hubiesen otras más. Ciertos rumores y especulaciones recientes ofrecen la esperanza de una explicación para estas peculiaridades; en ellos se sugiere que la distinción entre electrón y muón quizá no sea absoluta y también que quizás el muón y electrón no sean un par aislado de partículas, sino los primeros miembros de una numerosa familia.

Tanto el electrón como el muón son leptones: partículas sin un tamaño medible y que no dan señales de poseer una estructura interna. Los únicos otros leptones conocidos son dos tipos de neutrinos (uno para el electrón y otro para el muón). Además de carecer de todas las propiedades de que carecen el electrón y el muón, los neutrinos carecen asimismo de carga eléctrica y totalmente (o casi totalmente) de masa.

La masa del muón es más de 200 veces mayor que la del electrón, pero por lo demás ambas partículas son imprescindibles. Sin embargo, no son intercambiables: el muón no puede desprenderse simplemente de su masa extra y convertirse en un electrón; antes de lograrlo es preciso que se desprenda de cierta esencia indefinible (la "mu-idad").

Hasta el presente, la única modalidad confirmada de desintegración para el muón es su transformación en electrón mediante la emisión de un neutrino de tipo muónico (que lleva consigo la mu-idad) y de un antineutrino de tipo electrónico. Otras posibles modalidades de desintegración, ampliamente debatidas en meses recientes y tal vez observadas experimentalmente, implican una violación de la conservación de la mu-idad.

Un grupo de siete investigadores del Instituto Suizo de Investigación Nuclear (SIN), cercano a Zürich, informó de la posible observación de una desintegración anómala del muón. Con el ciclotrón de 590 millones de electronvoltios del SIN, detectaron media docena de sucesos (de entre varios miles de millones) en los que un muón parecía desintegrarse en un electrón y un rayo gamma, o fotón de alta energía. El rayo gamma puede llevar consigo la masa del muón, pero no la propiedad de mu-idad.

Los investigadores suizos postergaron la publicación o el anuncio público de estos hallazgos preliminares hasta que el experimento pudiese repetirse y pudiesen recogerse más datos. No obstante, la noticia acerca del posible descubrimiento suscitó especulaciones teóricas inmediatas sobre la significación y consecuencias del experimento. Ahora, al parecer, se está muy lejos de tener la certidumbre de que de hecho se hayan observado desintegraciones anómalas de muones en el SIN; no se ha logrado reproducir ese resultado hasta la fecha en otras partes, pero el interés en el tema sigue siendo grande. Los investigadores suizos no han publicado todavía un informe sobre lo que descubrieron, pero han comenzado a aparecer interpretaciones teóricas sobre tales hallazgos. Estas interpretaciones suponen unos ajustes radicales al modelo que los físicos han formado sobre el universo. Introducen la posibilidad de una nueva fuerza fundamental en la naturaleza, atribuyen masa a los neutrinos y postulan nuevos leptones.

Una de las interpretaciones es la de Steven Weinberg, de la Universidad de Harvard, y James D. Bjorken, del Stanford Linear Accelerator Center (SLAC). En un artículo aparecido en *Physical Review Letters,* estos investigadores postulan un modelo de los leptones en el cual la conservación de la mu-idad no es una ley básica de la naturaleza; se suscita, en cambio, como un rasgo casi accidental de una extensa clase de teorías físicas. Sucesos tales como el que un muón se desintegre formando un electrón y un rayo gamma no estarían prohibidos, pero estarían suprimidos; tales sucesos no podrían producirse a través de ninguna de las cuatro fuerzas básicas conocidas de la naturaleza (las fuerzas fuerte, electromagnética, débil y gravitacional). La observación de tales sucesos podría entonces implicar la existencia de una quinta fuerza, no afectada por la supresión, a la que se ha dado el nombre de "superdébil". Un aspecto interesante de este modelo consiste en que la supresión aparece solamente si al menos un tipo de neutrino posee una masa mayor que cero. Tal cosa es ciertamente posible; los límites experimentales superiores respecto de la masa de los neutrinos no son muy restrictivos.

Otro planteamiento del problema de las desintegraciones que cambian la mu-idad ha sido el propuesto por T. P. Cheng, de la Universidad de Missouri, y Ling-Fong Li, de la Universidad Carnegie-Mellon. Sus comentarios, publicados asimismo en *Physical Review Letters,* fueron redactados antes de que se conocieran los descubrimientos del SIN. Señalan que la probabilidad de sucesos tales como la desintegración de un muón en un electrón y un fotón aumentaría muchísimo si existiesen otros leptones. Entre tales nuevos leptones habría probablemente partículas de masa mucho mayor que la del electrón y del muón, y habría asimismo nuevos neutrinos.

Las pruebas que sugieren la posibilidad de leptones adicionales se han estado acumulando durante más de un año en experimentos realizados en varios laboratorios.

En el SLAC, por ejemplo, en los productos de colisiones entre electrones y positrones se ha observado una discrepancia pequeña, pero uniforme, que podría ser indicio de que se ha observado un nuevo leptón. La discrepancia consiste en algunos muones extra, que no es posible justificar fácilmente mediante otros mecanismos. No existe una interpretación convincente de estos sucesos, pero está ganando verosimilitud la hipótesis de que los muones provengan de la desintegración de un leptón eléctricamente cargado con una masa cuyo equivalente oscilaría entre 1600 y 2000 millones de electronvoltios. En otros laboratorios, experimentos muy diferentes han sugerido nuevos leptones con otras masas, incluyendo tal vez un leptón de gran masa y eléctricamente neutral.

Si resulta que los leptones constituyen una gran familia de partículas, el contarlos y catalogarlos podría conver-

tirse en una ocupación importante de los físicos a lo largo de la próxima década. A. K. Mann y Henry Primakoff, de la Universidad de Pennsylvania, han propuesto un experimento que podría proporcionar una valiosa pista inicial acerca de su número. El experimento podría asimismo señalar un nuevo límite superior para las masas de los neutrinos. Es un plan ambicioso de ciencia física a escala continental.

Mann y Primakoff proponen buscar unas "oscilaciones neutrínicas", conversiones espontáneas de un tipo de neutrino en otro. Dichas oscilaciones no son desintegraciones de una partícula que den por resultado la otra, sino cambios reversibles de estado mecánico-cuántico. En cierto sentido, son manifestaciones de una ambigüedad subyacente en la identidad de los neutrinos. Si las oscilaciones son posibles, un haz puro de neutrinos de tipo muónico podría llegar a contaminarse progresivamente con neutrinos del tipo electrónico. Sea cual fuere la composición original del haz, con el tiempo tiene que llegar a ser una mezcla igual de ambos tipos de neutrinos (o de todos los tipos de neutrinos, si hubiese más de dos). Las oscilaciones neutrínicas están previstas sólo si se satisfacen ciertas limitaciones respecto de la teoría de los leptones; en esa medida constituyen una comprobación de la teoría.

Las oscilaciones serían posibles sólo si la conservación de la mu-idad no fuese absoluta. En otras palabras, el neutrino de tipo muónico puede convertirse libremente en un neutrino de tipo electrónico sólo si un muón puede asimismo convertirse en un electrón. Otra condición consiste en que las masas de los neutrinos no pueden ser todas iguales a cero, y que si son mayores que cero, no pueden ser todas iguales.

Si pudiesen detectarse las oscilaciones neutrínicas, la proporción de cada tipo de neutrino en un haz que haya llegado al equilibrio proporcionaría una medida sencilla del número de tipos de neutrinos. Por ejemplo, si el tipo muónico compone la mitad del haz, existen sólo los dos tipos conocidos; si constituye un tercio, hay otra especie de neutrino no descubierto.

Como quiera que los neutrinos y los leptones de gran masa parecen darse en parejas, la proporción serviría también para estimar el número total de electrones.

La dificultad para medir oscilaciones neutrínicas es una dificultad de escala. Se prevé que las oscilaciones sean bastante

lentas, y que un neutrino, desplazándose prácticamente a la velocidad de la luz, recorrería una distancia considerable antes de que la probabilidad de una oscilación llegase a ser realmente significativa.

Una longitud apropiada para medir las oscilaciones sería de 1000 kilómetros. A causa de la distintiva carencia de propiedades del neutrino y de su renuencia a interaccionar con la materia ese grandioso experimento podría ser factible. Mann y Primakoff proponen dirigir un haz neutrínico de alta energía, como el generado en el Fermi National Accelerator Laboratory, hacia el suelo formando un pequeño ángulo. Los neutrinos pasarían entonces sin estorbo a través de la tierra sólida. A 1000 kilómetros de distancia, en donde el haz emergiera del suelo formando el mismo ángulo pequeño, se colocaría un detector que discriminase entre neutrinos de tipo muónico y neutrinos de tipo electrónico.

El láser de electrones libres

Reducido a su forma esencial, un láser es un oscilador electrónico: funciona induciendo movimientos idénticos en numerosísimos electrones. En un láser real, esa simplicidad subyacente queda oscurecida porque los electrones que emplea el oscilador son los que están ligados en la intricada estructura de átomos o moléculas. Ahora se ha llegado a construir y a hacer funcionar un láser en el cual se elimina esas complicaciones. Como medio emplea un haz de electrones al vacío.

El medio de un láser convencional consiste en una población de átomos o moléculas a los que, mediante una fuente externa de energía se puede "bombear" hasta un estado excitado. La acción láser comienza cuando algunos de los átomos o moléculas caen hacia un estado de menor energía al emitir un fotón, o cuanto de radiación electromagnética. Algunos de los fotones emitidos son atrapados en una "cavidad" óptica formada por un par de espejos, y conforme los fotones rebotan de ida y vuelta a través del medio del láser, estimulan a otros átomos o moléculas excitados a emitir radiación de la misma frecuencia y fase. La transmisión parcial efectuada por uno de los espejos permite que parte de la radiación se escape como un haz coherente.

Cuando un átomo absorbe o emite radiación, lo hace cambiando el estado de movimiento de un electrón ligado. Una de las premisas básicas de la mecánica cuántica es que los electrones de un átomo sólo pueden tener modos discretos de movimiento y, por consiguiente, los átomos sólo pueden poseer niveles discretos de energía; los estados intermedios están prohibidos. El átomo puede absorber o emitir sólo los fotones cuyas energías correspondan a la diferencia de energía existente entre dos estados permitidos.

A causa de las restricciones de los movimientos de los electrones en el interior de los átomos, la mayoría de los láseres poseen una sola longitud de onda característica que no se puede modificar fácilmente. Las excepciones más importantes las constituyen los láseres cuyo medio de operación es un colorante orgánico, una molécula provista de una serie de líneas de absorción y emisión tan cercanas entre sí que equivalen a un continuo. Los láseres de colorante pueden sintonizarse a cualquier longitud de onda comprendida dentro del continuo. Sin embargo, el espectro correspondiente a un determinado colorante es relativamente estrecho. Además, los láseres de colorante no pueden irradiar cantidades muy grandes de energía.

Un electrón libre no está sujeto a las mismas restricciones que un electrón situado en un átomo, y puede poseer prácticamente cualquier estado de movimiento y cualquier cantidad de energía. Como un electrón emite radiación siempre que cambia su estado de movimiento (es decir, siempre que se acelera), un electrón puede irradiar a cualquier longitud de onda. Esa capacidad de emisión infinitamente ajustable se aprovecha en muchos de los osciladores electrónicos empleados para generar radioondas y microondas. El láser de electrones libres podría proporcionar la misma capacidad de sintonización mediante el giro de una perilla a longitudes de onda ópticas.

El prototipo de láser de electrones libres fue construido hace poco por John M. Madey, de la Stanford University, y un grupo de colegas suyos, entre ellos David A. G. Deacon, Luis R. Elías, Gerald J. Ramian, H. Alan Schwettman y Todd I. Smith.

El corazón del láser es un imán superconductor de unos cinco metros de longitud enrollado en forma de doble hélice con una vuelta cada tres centímetros. El campo que éste produce tiene forma de tirabuzón. Los electrones proporcionados por el acelerador electrónico lineal de Standford se dirigen a través de una cámara de vacío situada dentro del enrollamiento magnético. La radiación es emitida cuando los electrones se topan con el campo magnético, el cual les obliga a seguir trayectorias helicoidales. Lo que primero se emite es una radiación espontánea y no la radiación estimulada, característica del funcionamiento de todo láser. La radiación estimulada aparece cuando alguno de los fotones emitidos espontáneamente son reflejados y vuelven a interactuar con los electrones en movimiento.

El acelerador proporciona los electrones en "paquetes" de alrededor de un milímetro de longitud, y el láser funciona solamente durante los breves intervalos en los que un paquete de electrones está situado dentro del campo magnético. Así pues, la emisión del láser está compuesta de un tren de breves impulsos luminosos, de alrededor de un milímetro de longitud cada uno. La longitud de onda de la radiación emitida está condicionada por la intensidad y la geometría del campo magnético y por la energía de los electrones, magnitudes que pueden estar bajo el control del experimentador.

En la máquina de Standford se demostró el funcionamiento del láser con un campo magnético de 2,4 kilogauss y con electrones cuya energía era de 43,5 millones de electronvoltios. La radiación producida por el láser tenía 34.170 angstroms de longitud de onda, o sea, estaba situada en la región infrarroja del espectro electromagnético. La potencia media de la radiación externa era de 0,36 vatios, pero esta pequeña magnitud refleja principalmente el hecho de que el láser estaba apagado la mayor parte del tiempo (entre uno y otro "paquete"). En el interior de la cavidad del láser, la intensidad máxima del campo de irradiación era de unos 500 kilovatios.

El motivo principal para el interés que despierta el láser de electrones libres, aparte su elegante demostración de principios físicos, reside en la comodidad con que se le puede ajustar a lo ancho de una amplia banda de longitudes de onda. La obtención de pequeñas longitudes de onda está limitada por la necesidad de una densidad creciente de "paquetes" de electrones, pero probablemente sería posible lograr longitudes de onda de no más de 1200 angstroms con una máquina que podría construirse con los conocimientos actuales. Dicha máquina sería una fuente de radiación monocromática intensa, que podría ajustarse a voluntad a prácticamente cualquier longitud de onda en la zona infrarroja, visible o ultravioleta del espectro.

La defensa en los opistobranquios

Los llamativos colores de estos moluscos gasterópodos no son un mero capricho de la naturaleza; hoy se empieza a comprender que están relacionados con sistemas defensivos muy variados y eficaces

Joandomènec Ros

En los fondos someros o profundos de los océanos de todo el mundo, ya sea en las aguas cálidas de los trópicos o en las frías de los mares templados o árticos, existe una enorme variedad de animales y de plantas pertenecientes a los grupos sistemáticos más diversos y todos perfectamente adaptados a un ambiente tan extraño y fascinante para el hombre como es el bentónico, es decir, el constituido por la frontera entre el sustrato sumergido y el agua que lo baña. Las cuencas oceánicas han sufrido en el transcurso de los tiempos avatares de toda índole, pero sus condiciones ambientales han variado muy poco, en general, a lo largo de la historia de la Tierra, y ello ha permitido la conservación y la evolución en un ambiente muy estable de estirpes enteras de organismos que, en su mayoría, combinan a la perfección un cúmulo de caracteres arcaicos con otros muy especializados.

Cualquiera de estos grupos animales o vegetales merece un estudio profundo, pues conocer su biología puede tener implicaciones de la mayor importancia, inmediata o futura, para el hombre, que en un mundo superpoblado se está planteando cada vez más seriamente la posibilidad de explotar todos los recursos utilizables. Se comprende que la biología del mejillón, de la merluza o de las algas unicelulares del plancton, por ejemplo, tenga algo más que un interés puramente científico cuando se trata de aumentar las fuentes de proteínas existentes y de hallar otras nuevas. Pero no sólo las especies animales y vegetales de interés comercial pueblan el mar; las comunidades marinas están formadas por una variedad enorme de organismos, todos implicados en un proceso continuo de interacciones mutuas que posibilitan la vida oceánica, y todos unidos por relaciones de dependencia directa o secundaria que hacen que el último de sus representantes sea un nudo fundamental en cualquiera de las innumerables y extraordinariamente complejas redes tróficas que mantienen la vida marina.

Entre estos animales poco conocidos, de utilidad aparentemente nula para el hombre, se encuentran unos pequeños moluscos, los opistobranquios, que constituyen una subclase de gasterópodos (otras dos subclases nos resultan más familiares: los prosobranquios, que son los típicos caracoles marinos, y los pulmonados, que agrupan a los caracoles y babosas terrestres y de las aguas dulces). Los opistobranquios abarcan especies enormemente variadas, casi exclusivamente marinas y en su mayoría bentónicas. Sus principales características, además de las que comparten con todos los gasterópodos, derivan de la liberación de las limitaciones estructurales, funcionales y de comportamiento que suponen la concha y la masa visceral típicamente arrolladas y asimétricas. En efecto, los opistobranquios, y en especial los grupos más evolucionados, se han emancipado total o parcialmente de la concha típica de los moluscos y se han convertido en animales desnudos (por lo que a veces se les llama también babosas de mar). En los opistobranquios más primitivos la cavidad paleal (situada entre el manto y la concha) se ha reducido, y falta totalmente en los más avanzados; la branquia original, o ctenidio, se mantiene en algunos grupos, pero desaparece junto con la cavidad paleal, como ocurre con el osfradio u organo olfatorio. Para desempeñar sus funciones se desarrollan órganos respiratorios, principalmente dorsales (branquias), y olfatorios, cefálicos (rinóforos), de neoformación.

El cuerpo entero, libre de la tiranía que imponía la concha, adopta formas muy variadas, secundariamente simétricas, y, en casi todas las especies, coloraciones vistosas de significados muy diversos, algunos de los cuales se analizarán aquí. El aparato bucal y la rádula (la lengua córnea de los gasterópodos) sufren pequeñas transformaciones, derivadas de su adaptación a un amplio espectro de regímenes alimentarios que suelen ser muy especializados y explican muchos rasgos de la biología y de la ecología de estos animales. Todos ellos son hermafroditas, de fecundación cruzada, y depositan puestas bentónicas, raramente pelágicas, que dan origen a larvas velígeras libres y planctónieas o bien directamente a juveniles bentónicos bastante desarrollados. La vida de estos animales suele ser de alrededor de un año, alcanzándose los dos o incluso más en algunas especies grandes; todos son de dimensiones reducidas, de menos de 1 cm a los 10 cm como un máximo que sólo sobrepasan unas pocas especies.

Los opistobranquios son polifiléticos, es decir, los distintos órdenes que los integran no se hallan relacionados por un origen evolutivo próximo común. En los distintos órdenes es posible hallar formas completamente desnudas, en otros la reducción de la concha, o la presencia de papilas dorsales, u otros caracteres, han aparecido de manera independiente y a través de una evolución convergente. Veamos cuáles son estos grupos taxonómicos.

Los opistobranquios testáceos (con concha) o tectibranquios (con la branquia cubierta) son los más primitivos, y pertenecen principalmente al orden cefalaspídeos; la concha suele ser pequeña y el molusco no puede retirarse por completo a su interior, debido sobre todo al amplio pie, propio de animales excavadores. El aspecto general es el de un típico "caracol" marino, salvo en las especies cuya concha es parcialmente interna; las formas más primitivas son

		ESTRATEGIAS DEFENSIVAS			
		DEFENSAS PRIMARIAS O INDIRECTAS		DEFENSAS SECUNDARIAS O DIRECTAS	
		Pasar inadvertido al depredador	Resultar "no apetecible" al depredador	Huir y/o desconcertar al depredador	Contraatacar al depredador
MECANISMOS DEFENSIVOS	ESTRUCTURALES	Cripsis: homocromía (pigmentos propios) Cripsis: homotipia Cripsis: coloración disruptiva Cripsis: contrasombreado Cripsis: eliminación de la sombra lateral (?) y de la silueta	Concha, pseudoconcha, tegumento resistente Espículas Glándulas exocrinas, sacos cnidógenos y vacuolas epidérmicas Aposematismo Mimetismo	Glándulas exocrinas, sacos cnidógenos y vacuolas epidérmicas Autotomía	Glándulas exocrinas, sacos cnidógenos y vacuolas epidérmicas
	QUIMICOS	Secreciones de inmunidad o anestésicas	Secreciones venenosas Secreciones ácidas Secreciones repelentes Mal sabor generalizado (?) Secreciones de inmunidad o anestésicas	Secreciones de inmunidad o anestésicas Secreciones de enmascaramiento	Secreciones venenosas Secreciones ácidas Secreciones repelentes Mal sabor generalizado (?) Secreciones de inmunidad o anestésicas
	DE COMPORTAMIENTO	Anacoresis Anacoresis trófica	Exhibición individual y tanatosis Exhibición de grupo	Recciones de huida y/o de evitación Exhibición individual y tanatosis Comportamiento deimático	Defensa agresiva
	CLEPTODEFENSAS	Cripsis: homocromía (pigmentos adquiridos) Cripsis: epibiontes	Cleptocnidos Secreciones defensivas y mal sabor (?)	Secreciones defensivas y mal sabor (?)	Cleptocnidos Secreciones defensivas y mal sabor (?)

LOS MECANISMOS DEFENSIVOS de los opistobranquios se disponen en este cuadro según las estrategias de defensa en las que se hallan integrados. En realidad, estos mecanismos que aquí aparecen individualizados para mayor claridad, en la naturaleza no suelen presentarse por separado, y con frecuencia dos o más de ellos forman parte del comportamiento protector de una especie determinada, tanto más cuanto que algunos son consecuencia de otros (las coloraciones aposemáticas, por ejemplo, se presentan siempre en especies bien defendidas –por uno o más de los sistemas considerados– o en especies que las imitan). Algunos mecanismos son de presencia dudosa o no hay evidencia suficiente que apoye su existencia, lo que se indica con un interrogante. Los mecanismos defensivos se han agrupado según su naturaleza en estructurales, químicos, derivados del comportamiento (esencialmente motores y disparadores de los demás) y mecanismos que emplean defensas robadas (cleptodefensas). Las estrategias defensivas se califican de primarias o indirectas si contribuyen a que un depredador no inicie un intento de captura de la presa, o de secundarias o directas si actúan después de que el depredador haya iniciado un intento de captura. En el primer caso el animal presa es ignorado; en el segundo, el sistema de defensa posibilita que aquél escape a la depredación.

detritívoras, y macrófagas y cazadoras las más evolucionadas.

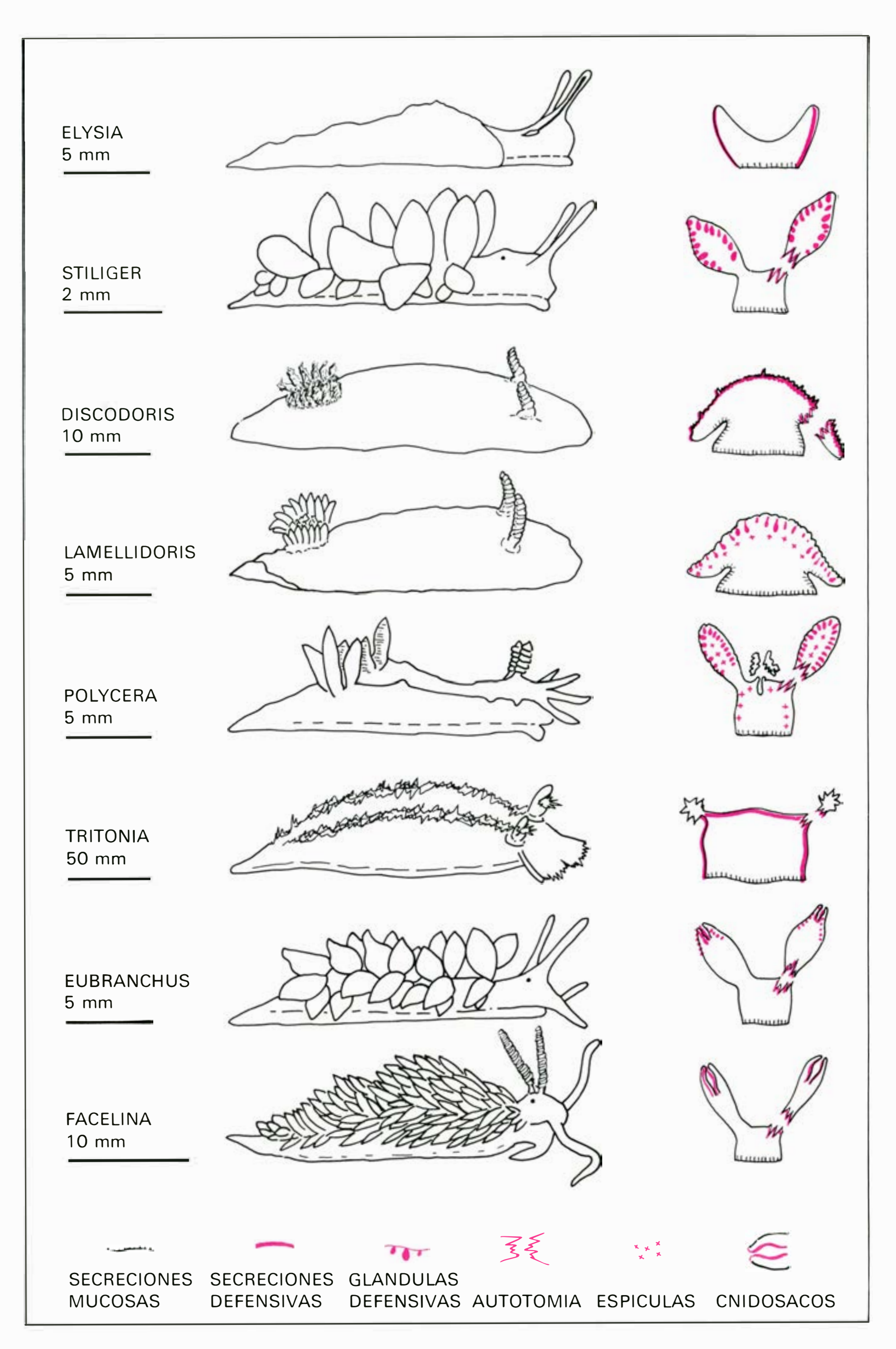
Los piramideláceos tienen un aspecto externo típicamente prosobranquio (hasta hace poco tiempo se les clasificaba en este grupo), pero son ectoparásitos de animales bentónicos y suctores de sus jugos internos. Los acoclidiáceos, muy pequeños, están adaptados a su tamaño y a la vida intersticial (entre los granos de arena de los fondos marinos). La mayoría de especies carecen de concha y en su lugar se desarrollan espículas calcáreas, que encontraremos también en otros grupos.

Los aplisiáceos constituyen un orden bastante primitivo de opistobranquios; no son excavadores (de hecho, algunas especies nadan desgarbadamente mediante los movimientos de los amplios parápodos en que se convierte el manto), su cuerpo es alto y la concha, vestigial, está reducida a una placa córnea oculta parcialmente en la cavidad paleal. Macroherbívoros, estos animales se alimentan mordisqueando los frondes de gran variedad de algas; son las conocidas liebres de mar, de tamaño en ocasiones considerable.

El cuerpo de los pleurobrancáceos es más aplastado, la cavidad paleal ha desaparecido, muy pocas especies conservan la concha (en todo caso reducida) y algunas son nadadoras; su régimen alimentario, sedimentívoro o suctor de jugos de esponjas y ascidias, los asemeja a los verdaderos nudibranquios, a los que seguramente dieron origen. Aunque diversas especies de los distintos grupos son pelágicas (viven flotando o nadando en las aguas libres y no sobre el fondo), dos órdenes de opistobranquios lo son enteramente: los tecosomas, con concha y micrófagos, y los gimnosomas, sin ella y depredadores, sobre todo de tecosomas. Ambos grupos forman una parte importante del zooplancton marino y son conocidos colectivamente como pterópodos.

En los nudibranquios (de branquias desnudas), la concha, la cavidad paleal y la branquia primitiva se pierden en el adulto, y la superficie dorsal desnuda queda libre para desempeñar las funciones sensoriales, respiratorias y defensivas necesarias. Los dendronotáceos son nudibranquios primitivos, aunque especializados en determinados aspectos; el ano se abre todavía a la derecha del cuerpo y los rinóforos no son retráctiles. Se alimentan de cnidarios sésiles, aunque algunas formas son macrófagas y depredadoras. Este hábito lo han conservado, diversificándolo, los eolidáceos, que son gasterópodos de cuerpo esbelto que almacenan los nematocistos (cnidos, cápsulas provistas de filamentos urticantes) de sus presas en cnidosacos de sus apéndices dorsales o *cerata*, desde donde pueden expulsarlos para su propia defensa. Su coloración suele ser vistosa, ya sea a nivel de todo el cuerpo o limitada a los apéndices dorsales.

En los doridáceos el cuerpo es generalmente oval y aplanado, aunque muchas especies presentan excrecencias dorsales de forma diversa; el ano se ha desplazado a la parte posterior del dorso y está rodeado por un penacho de branquias de neoformación; branquias y rinóforos son retráctiles. El manto suele presentar espículas y, por lo general, se halla adornado de colores muy variados. Estos animales son microcarnívoros: ramonean esponjas, briozoos, ascidias y otros invertebrados sésiles.



LOS MECANISMOS DEFENSIVOS DE LOS OPISTOBRANQUIOS se han esquematizado en esta figura. En los cortes transversales a la derecha se indica la presencia de distintos atributos defensivos (en rojo: glándulas, espículas, cnidosacos, autotomía, etc.) en unos pocos géneros representativos de sacoglosos *(Elysia* **y** *Stiliger),* **doridáceos** *(Discodoris, Lamellidoris, Polycera),* **dendronotáceos** *(Tritonia)* **y eolidáceos** *(Eubranchus, Facelina).* **Varios de estos mecanismos pueden encontrarse reunidos en un mismo animal, que asimismo puede realzarlos u ocultarlos mediante los distintos tipos de coloraciones adaptativas (aposematismo, cripsis) señaladas en el texto.**

Los armináceos, el último grupo de verdaderos nudibranquios, están modificados para la vida excavadora; algunos poseen formas aberrantes y su alimentación es enteramente suctorial.

Los sacoglosos no están relacionados directamente con los demás nudibranquios; se caracterizan sobre todo por alimentarse de los líquidos celulares de algas marinas, que succionan después de perforar las paredes de las mismas. Existen formas desnudas y testáceas (entre ellas los únicos gasterópodos "bivalvos"), limaciformes y eolidiformes, casi siempre de tamaño reducido debido al tipo especial de alimentación, que también ha facilitado en algunas especies la existencia de simbiosis con algas unicelulares y con cloroplastos de algas pluricelulares, que suelen dar a estos moluscos su color verde.

Los opistobranquios, en especial los nudibranquios, ocupan nichos ecológicos muy concretos que rara vez les disputan las especies de otros grupos de animales marinos. Su régimen alimentario, casi siempre ramoneador (de algas, esponjas, cnidarios, briozoos, etc.) y muy especializado (la mayoría de ellos se alimentan sólo de una o de muy pocas especies de presas), restringe su distribución a la de la especie alimento, lo que da origen a asociaciones depredador-presa muy estrictas, tanto en el tiempo como en el espacio, a la vez que segrega a cada especie de opistobranquio en un nicho ecológico preciso, por lo que la competencia por el alimento es muy reducida. Las diversas comunidades bentónicas marinas (los fondos de fango, de arena, los algares, las praderas de fanerógamas marinas, los sustratos rocosos, los coralígenos, etc.) tienen todas su fauna de opistobranquios propia, y en cada comunidad estos moluscos representan un elemento faunístico característico.

Esta breve introducción en torno a la sistemática y a la biología de los opistobranquios no estaría completa si no se hiciera referencia al hecho de que, en comparación con los demás moluscos, estos animales son todavía mal conocidos. La falta de concha, sobre todo, así como el hecho de que pocas especies se presenten en poblaciones numerosas, han contribuido a que, durante muchos años, los investigadores relegaran el estudio de estos gasterópodos; sobre todo los conquiliólogos, a quienes tanto debe la malacología, prefieren en general el estudio de la concha.

El panorama está cambiando, afortunadamente, y hoy día son numerosos los estudiosos, que, en todo el mundo, se esfuerzan en poner al día la sistemática de estos animales y en dilucidar los todavía numerosos problemas que plantean su anatomía, su biología y su ecología.

Hace ahora más de quince años vi un opistobranquio por primera vez: se trataba de una liebre de mar que se movía de manera indolente sobre un limpio fondo de arena de una playa mediterránea. Conservo todavía viva su imagen; desde entonces, y aunque los primeros descubrimientos fortuitos y gozosos fueron sustituidos más tarde por la observación detallada y el estudio científico, en todas mis inmersiones no deja de maravillarme el magnífico colorido, la delicada estructura y la gran variedad de formas de estos animalitos frágiles y, a la vez, tremendamente eficaces cuando se trata de asegurar su supervivencia.

¿Qué razones explican esta riqueza cromática en animales que, a diferencia de las aves, por ejemplo, no basan en la visión su estrategia de lucha por la existencia? ¿Qué ventajas supone para estos moluscos la posesión de coloreadas libreas, libreas y colores que la selección natural ha conservado y, por tanto, es de suponer que "sirven" para algo? Parece claro hoy día que, descartados para estos gasterópodos los usos que otros animales hacen de sus coloraciones (sexualidad, territorialismo, etc.), hay que buscar en la defensa frente a los depredadores la utilidad primera de las coloraciones y de los diseños de color de los opistobranquios.

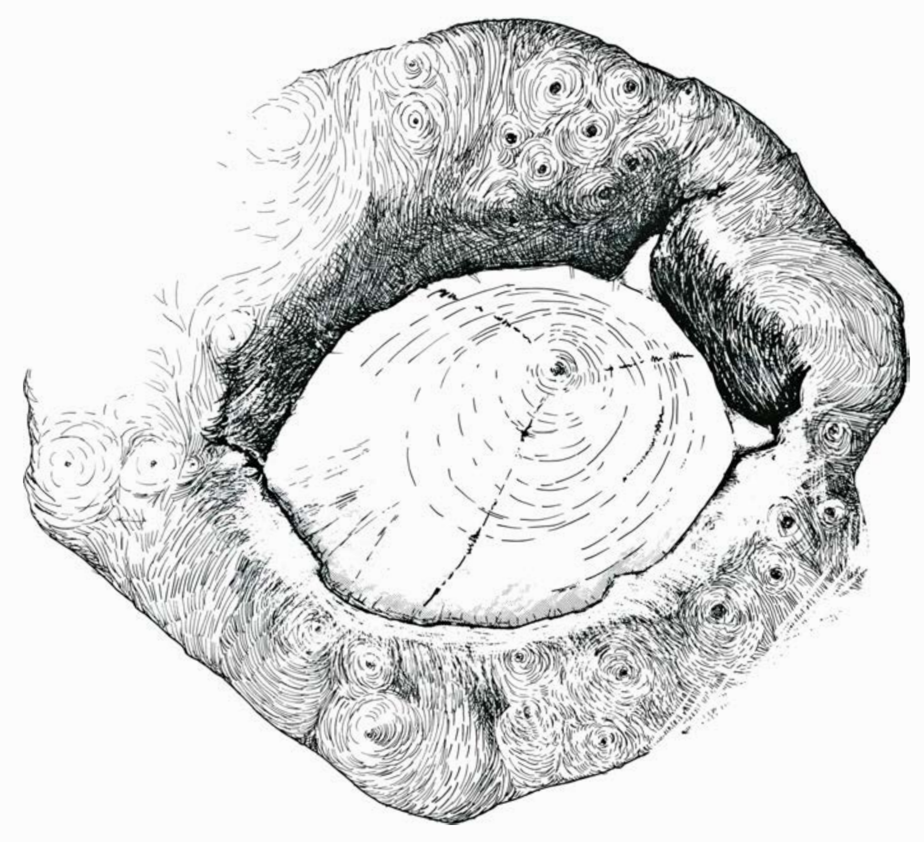
Las coloraciones vivas, a menudo uniformes, otras veces abigarradas o contrastantes, suelen obedecer casi siempre a dos tipos de estrategia defensiva: confunden al animal con su entorno o sustrato, con lo que aquél pasa inadvertido a los depredadores; o, por el contrario, hacen que el animal aparezca nítido, destacando de su entorno, para que un depredador potencial asocie los colores llamativos con algún carácter defensivo, ya sea éste la posesión de armas especializadas o de un sabor desagradable. En el primer caso se habla de coloraciones crípticas o que enmascaran, en el segundo de coloraciones aposemáticas o que advierten. Por lo menos parte de estas coloraciones (la mayoría de las advertidoras) se fundamentan en la existencia de algún tipo de defensa antidepredador, mientras que la cripsis no precisa de una base defensiva previa; en la práctica, existen animales crípticos defendidos por otros mecanismos, y animales aposemáticos no defendidos; en este último caso se habla de mimetismo. La utilización del color con los fines descritos es común en los opistobranquios.

Como introducción a la utilización de las coloraciones adaptativas por parte de estos animales será adecuado comentar brevemente los numerosos y diversos caracteres que los opistobranquios utilizan en sus estrategias de supervivencia. El autor ha intentado sistematizarlos de algún modo y ha distinguido entre defensas estructurales, químicas, de comportamiento y cleptodefensas (defensas que se toman prestadas de otros organismos), atendiendo a la naturaleza de los mecanismos defensivos, o bien, según cómo se utilicen éstos, entre defensas primarias o indirectas (las que contribuyen a que el depredador no inicie el intento de captura de la presa, que es ignorada) y secundarias o directas (las que funcionan después de producirse un intento de captura, con lo que posibilitan que ésta no se materialice). En la tabla de la página 49 se recogen estos distintos mecanismos; aunque aquí se han etiquetado y separado para mayor claridad, no sólo no suelen presentarse aislados, sino que con frecuencia la existencia de uno de ellos implica la de otro u otros, y es común por tanto que dos o más de ellos se encuentren en una misma especie.

Los ejemplos que se mencionan en el texto corresponden casi exclusivamente a especies de aguas ibéricas, mediterráneas y atlánticas, con las que el autor está más familiarizado, pero debe señalarse que para muchas otras especies de opistobranquios de todo el mundo se han descrito mecanismos parecidos. Ello no implica que las interpretaciones que aquí se hacen de ciertas estructuras y comportamientos sean compartidas por todos los investigadores de este grupo; algunas coloraciones aposemáticas y el mimetismo, en particular, se basan en datos todavía fragmentarios e incompletos, lo que hace que el escepticismo de determinados autores esté en parte justificado.

Ya he indicado que la concha de los tectibranquios es reducida, casi siempre sin opérculo, y que los nudibranquios carecen de ella; para muchos moluscos de todos los grupos encerrarse dentro de la concha resulta una defensa eficaz contra depredadores no demasiado avispados, pero los opistobranquios no pueden confiar su defensa a un abrigo que, en el mejor de los casos, sólo en parte los cobija. A este propósito se ha dicho que, en estos moluscos, la defensa pasa de ser mecánica y pasiva a ser química y biológica y activa; ello es cierto, como veremos a continuación.

Algunas especies de opistobranquios tienen el cuerpo desnudo coriáceo, por lo menos el dorso, pero ello es la excepción, pues el cuerpo de estos moluscos suele ser blando. En muchas especies, sin embargo, queda asegurada cierta rigidez del manto por multitud de espículas calcáreas embebidas, de manera regular o no, en el tegumento dorsal, y a veces también en el pie y en los órganos sensoriales. Estas espículas, normalmente alargadas, pueden ser vestigios de la concha ancestral o haberse originado secundariamente como respuesta a la falta de estructuras duras en el cuerpo de estos animales. Muchos doridáceos (como *Peltodoris atromaculata* y *Polycera quadrilineata,* por ejemplo) presentan espículas, y en algunas especies éstas sobresalen del tegumento como dardos acerados recubiertos por excrecencias dorsales. La armadura que estas espículas suponen es suficiente para que algunos



TYLODINA PERVERSA excava su propio escondrijo en la masa carnosa de la esponja *Verongia aerophoba*, de la que posee el mismo color amarillo intenso (homocromía alimentaria). El ocultamiento es notable, y la forma de lapa del pleurobrancáceo dificulta su extracción. El autor denomina anacoresis trófica a este método de ocultación, propio de otras especies de opistobranquios.

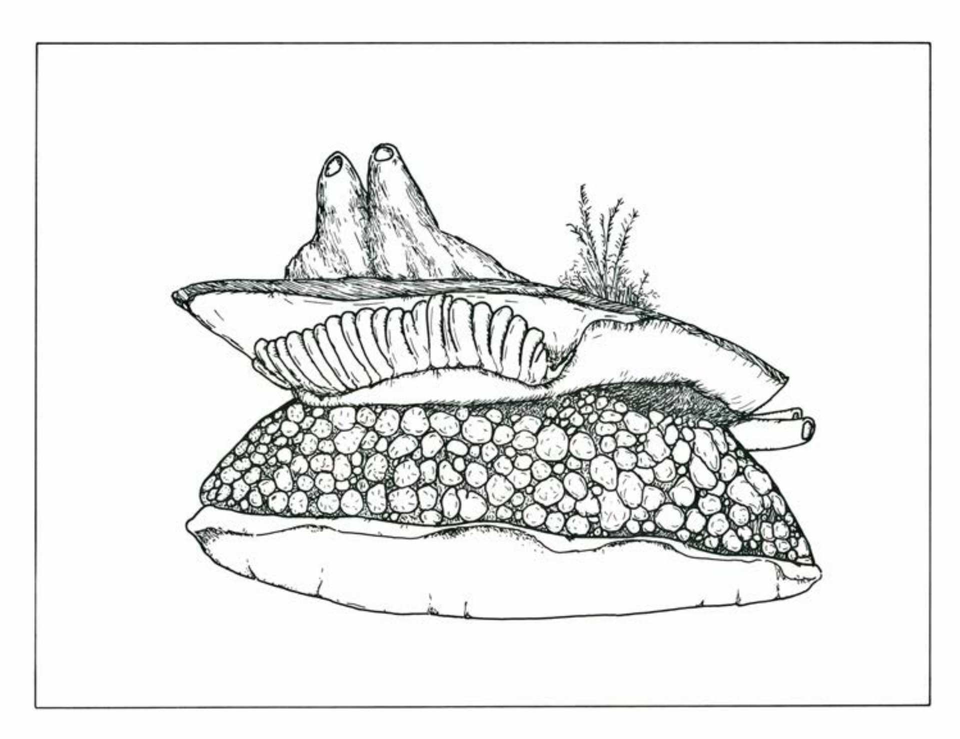
depredadores (otros moluscos carnívoros y peces) rechacen como alimento a los opistobranquios así protegidos.

Muchas especies de este grupo de gasterópodos secretan sustancias diversas, en general claramente relacionadas con la defensa. Algunas de estas secreciones son tóxicas y causan la muerte cuando se administran a peces, aún a concentraciones muy bajas, pero otras son de carácter típicamente repelente. Entre ellas destacan las secreciones de ácido sulfúrico muy concentrado (pH de 1 y 2) que expulsan algunos cefalaspídeos *(Philine)*, pleurobrancáceos *(Berthella, Oscanius, Pleurobranchus, Pleurobranchaea)* y doridáceos *(Doris, Lamellidoris, Discodoris, Anisodoris)*. Las glándulas exocrinas que producen estas secreciones derivan en la mayoría de los casos de glándulas mucosas que, en su origen, pudieron tener otras funciones (lubricantes, por ejemplo), y suelen hallarse localizadas en aquellas partes del cuerpo con las que es más probable que un depredador entre primero en contacto (el dorso y sus apéndices, respiratorios o sensoriales, que en ocasiones, además, son autotomizables y/o poseen una coloración vistosa que contrasta con la del resto del cuerpo del animal). La eficacia de una secreción ácida de este tipo debe quedar restringida, sin embargo, a un entorno mínimo, como la boca del posible depredador, pues en plena agua el ácido no sólo se diluye sino que se neutraliza rápidamente por la acción tamponadora del agua de mar. El mucus que suele secretarse abundantemente cuando existe una excitación táctil puede servir de vehículo a la secreción ácida y evitar que ésta se diluya en exceso. Esta defensa, como la presencia de espículas, obedece a una estrategia secundaria, mientras que otras que se comenterán a continuación entran dentro de la categoría de primarias.

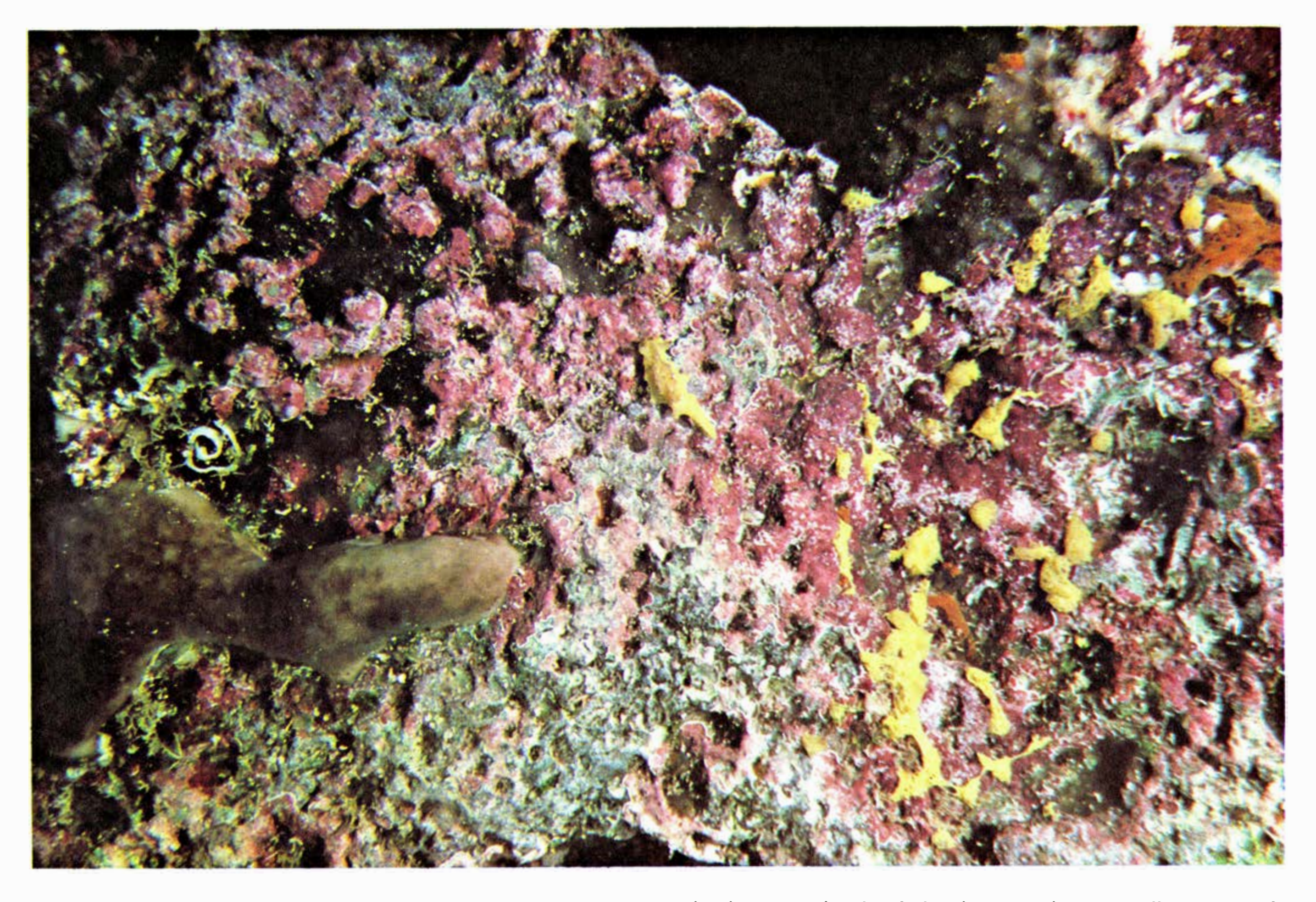
Se habla, en general, de secreciones repelentes para sistematizar de algún modo aquellos tipos de sustancias no ácidas que los opistobranquios segregan en presencia de un depredador; parecen proceder de glándulas no mucosas, epidérmicas o subepidérmicas, y se encuentran en prácticamente todos los grupos, en localizaciones no muy distintas a las comentadas a propósito de las glándulas de ácido. Estas secreciones comprenden materiales blancos, generalmente proteináceos y mucopolisacarídicos, que se exudan sólo cuando el animal es molestado y que, en la mayoría de los casos, aparecen como una tenue nube a su alrededor; se comprende que si en algunos casos estas sustancias son incoloras su detección se hace más difícil, pudiendo, sin embargo, ser más generales de lo que hasta ahora se supone. En otras ocasiones no existe ninguna secreción aparente, pero los animales son rehusados como alimento por depredadores experimentales; se habla entonces de "mal sabor", término que encierra más bien la ignorancia acerca de la naturaleza de la defensa implicada. Lo cierto es que peces de distintas especies mantenidos en acuario y hambrientos no se comen a los opistobranquios vivos que se les ofrecen, mientras que aceptan individuos muertos y, por tanto, incapaces de secretar sustancias repelentes.

Otras secreciones defensivas no se utilizan contra posibles depredadores, sino contra las presas. Los eolidáceos, cuyo régimen alimentario, como se ha indicado, incluye animales bien defendidos (los cnidarios poseen nematocistos o cnidos, cápsulas provistas de filamentos urticantes que utilizan para su propia defensa), deben solucionar dos problemas básicos relacionados con su alimentación: en primer lugar, inutilizar de algún modo las defensas de la presa; en segundo lugar, asegurar que las cápsulas urticantes que son ingeridas con el alimento no se disparen en su tránsito por el tubo digestivo, hacia el exterior (con las heces) o hacia los apéndices dorsales, donde serán almacenadas para su reutilización. El primer problema parece haberlo solucionado la evolución (al igual que en el caso de otros animales, como los peces comensales de anémonas de mar) mediante secreciones anestésicas o de inmunidad, mal conocidas, con las que el eolidáceo embadurna la presa antes de comerla, y que bien pudiera ser una simple barrera mucosa al disparo de los cnidos. (Otra posibilidad, no comprobada, sugiere que sería el eolidáceo el que se embadurnaría con el mucus del cnidario, de manera que utilizaría una especie de "disfraz" químico que la presa no distinguiría de sus propias secreciones). El segundo aspecto comentado parece estar resuelto, al menos en parte, por la existencia de numerosas vesículas características en el epitelio del tubo digestivo, así como por las mismas secreciones mucosas que al englobar a los cnidos deben aislarlos de los cambios de presión osmótica del tubo digestivo del eolidáceo, que podrían provocar su disparo en el interior del mismo; muchos cnidos que se expulsan con las heces, sin embargo, se han descargado.

Otro tipo de secreciones en los opistobranquios son propias de los aplisiáceos. De las tres especies de *Aplysia* de nuestras costas peninsulares, *A. depilans* excreta un líquido incoloro, lechoso, de olor penetrante y, según parece, tóxico; *A. punctata* segrega, además de éste, otro purpúreo formado por bilinas vegetales de color violeta y rojo, mientras que



SOBRE LA REDUCIDA CONCHA de este otro pleurobrancáceo *(Umbraculum mediterraneum)* crecen en ocasiones vegetales y animales sésiles propios de los sustratos sobre los que vive esta especie; el efecto críptico de los epibiontes hace prácticamente indistinguible al molusco de su entorno. En la figura aparecen una ascidia, hidrarios y algas. Briozoos, bivalvos y otros numerosos organismos forman parte importante, asimismo, en la configuración de este disfraz críptico.

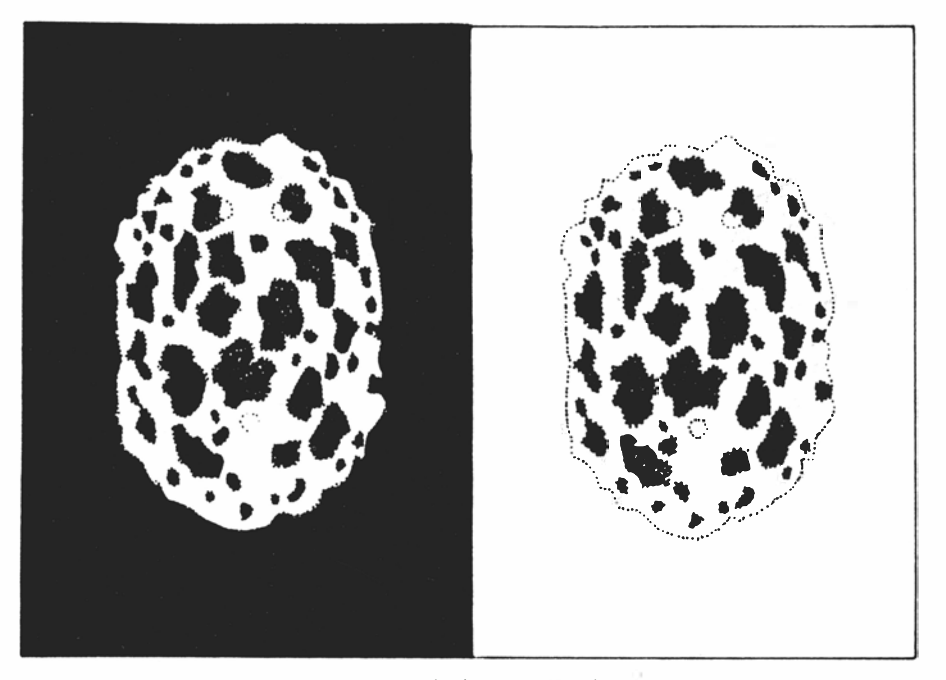


EL COLOR Y LA FORMA del doridáceo *Serigea sublaevis (en el centro)* **se combinan para conseguir un elevado efecto críptico sobre la esponja amarilla** *Clathrina coriacea;* **en esta fotografía es imposible distinguir a otro ejemplar de la misma especie que se halla enmascarado perfectamente sobre la esponja** *(abajo, derecha).* **A la izquierda, puesta de este nudibranquio de aguas canarias (fondos someros de Fuerteventura).**

A. fasciata exuda sólo las bilinas, que no son tóxicas, pero parecen proporcionar el tipo de defensa de "cortina de humo" tan común en los cefalópodos (en el pulpo y la jibia, entre otros, la secreción de la glándula de la tinta parece tener por misión que la atención del depredador se transfiera de la presa a un objeto inanimado, la nube de tinta, mientras aquélla escapa después de sufrir cambios de coloración que contribuyen a su enmascaramiento). Como sea que los dos tipos de secreciones se dan, separada o conjuntamente, en todas las especies de *Aplysia,* parece que podría tratarse de un doble sistema de enmascaramiento, visual y químico, que actuarían por separado (en *A. fasciata* y *A. depilans*) o al unísono (en *A. punctata*). Esta especie, así como otras liebres de mar no ibéricas que segregan a la vez los dos tipos de sustancias, deben poseer así una defensa más eficaz, pues en teoría son capaces de desorientar a un tiempo a los depredadores que cazan mediante la vista y a los que lo hacen valiéndose del olfato (ello en el caso de que la sustancia olorosa actúe como un "señuelo" químico, pero no se sabe si, de hecho, no es más que otro tipo de sustancia repelente).

Otro tipo de defensas no implican la utilización de ninguna estructura especial, sino la puesta en práctica de un determinado comportamiento, comportamiento, que, por otra parte, acompaña al empleo de otras estrategias defensivas complejas. Muchas especies de opistobranquios viven escondidas bajo piedras, enterradas en el sedimiento o en cualquier otro tipo de guarida; son, en la terminología de la defensa, anacoretas. Este comportamiento no es exclusivamente defensivo, sino que suele estar ligado a la búsqueda del alimento, y muchos doridáceos encuentran las esponjas o los briozoos de que se alimentan en los lugares más recónditos. En algunos casos la propia alimentación conduce a la construcción de una guarida muy especial. *Tylodina perversa* es un pleurobrancáceo que acostumbra a vivir en el interior de cavidades someras excavadas, "a mordiscos", en la matriz de la esponja *Verongia aerophoba.* Su forma de lapa hace en extremo difícil desalojarlo de estas cavidades y, además, otros dos factores contribuyen a su enmascaramiento: las chimeneas excavadas no son muy distintas, por el tamaño y la forma, de los ósculos de la esponja, y el color de *Tylodina* es amarillo vivo, exactamente igual al de su alimento. *Platydoris argo,* un doridáceo, se guarece de manera similar entre los encajes del briozoo *Sertella cellulosa,* de donde no suele ser posible extraerlo sin romper la estructura del briozoo. Otro pleurobrancáceo, muy similar a *Tylodina* pero de mayor tamaño y de colores apagados, *Umbraculum mediterraneum,* invierte el proceso de camuflaje: en lugar de hundirse en el sustrato éste crece sobre el molusco. Algas, hidrarios, lamelibranquios, ascidias y otros pobladores sésiles del bentos se instalan sobre el sombrero chino que es la reducida concha de este animal y ayudan a su enmascaramiento cuando el molusco se encuentra sobre un fondo rocoso cubierto de organismos incrustantes similares a los que él porta.

Para que esta defensa sea eficaz requiere una inmovilidad acentuada; todos



LA COLORACION DISRUPTIVA de *Peltodoris atromaculata* **hace que el animal sea críptico sobre su sustrato natural, la esponja** *Petrosia ficiformis,* **tanto si ésta es parda, color propio de sus algas simbiontes** *(izquierda),* **como si es blanca por vivir en lugar oscuro y carecer de algas** *(derecha).*

los animales crípticos suelen ser más o menos sedentarios y en ocasiones permanecen en el mismo lugar durante mucho tiempo, dificultando así la posible identificación por parte de los depredadores. (Algunas especies de opistobranquios emplean una inmovilización "cataléptica" cuando son molestados –tanatosis–, pero su utilidad parece ser la de desconcertar al depredador.) También los animales aposemáticos, como se verá, son de movimientos lentos, pero por razones muy distintas, mientras que otras especies de opistobranquios, se hallen o no defendidas, son capaces de huir ante una amenaza mediante movimientos rápidos, que en ocasiones producen una natación desgarbada (*Aplysia, Notarchus, Gasteropteron, Tritonia, Dendronotus,* etc.).

Acabamos de introducir algunos ejemplos de cripsis y será conveniente que nos extendamos un poco sobre este sistema defensivo de los opistobranquios. De la definición que se ha dado más arriba se advierte que existen diversos sistemas por los que un animal se confunde con su entorno. Se ha citado la homocromía (mismo color) de *Tylodina* y *Verongia,* y debe señalarse de entrada que una coloración igual o similar a la del sustrato es el tipo más común de cripsis en los opistobranquios, y ello por dos razones muy simples: el sustrato suele ser el alimento habitual de estos animales sedentarios, y muchos de ellos tienen el cuerpo translúcido, con lo que el pigmento de la ingesta colorea el cuerpo. La homocromía alimentaria es general en los sacoglosos (de colores verdes, pardos o rojos según las algas que coman), que además pueden acumular los pigmentos en forma de orgánulos (cloroplastos), útiles para el animal por otros motivos (simbiosis). Se presenta también extensamente en los eolidáceos, algunos de los cuales *(Flabellina affinis,* por ejemplo) efectúan puestas también homócromas, de modo que el progenitor las defiende en cierta manera, y, con menos frecuencia, en aplisiáceos (sobre todo en los juveniles), pleurobrancáceos, doridáceos y dendronotáceos. Muchas variedades, subespecies e incluso especies distintas de opistobranquios se han descrito en base a la coloración, pero se comprende rápidamente que ello carece de sentido en aquellas especies que pueden alimentarse de presas distintas cuyos colores son también diferentes. *Spurilla neapolitana,* que en nuestras aguas se alimenta indistintamente de los actiniarios *Aiptasia mutabilis* (pardoamarillenta), *Actinia equina* (verde o roja), *Anemonia sulcata* (verde o gris) y *Actinothoe sphyrodeta* (blancuzca); o *Favorinus branchialis,* que succiona el contenido de huevos de peces de distintos colores, son eolidáceos con un elevado "polimorfismo" cromático de origen trófico.

En muchas ocasiones, sin embargo, y ello supone una evolución conjunta depredador-presa (o, si se quiere, opistobranquio-alimento) mucho más prolongada, el color del molusco se debe a pigmentos propios, que a lo largo del tiempo la selección natural ha ido perfeccionando y adaptando a los de la presa. *Archidoris tuberculata,* por ejemplo, es un doridáceo cuya coloración parduscа abigarrada lo enmascara perfectamente cuando se halla sobre *Halichondria panicea,* la esponja que consume, y una subespecie roja de este doridáceo se alimenta de la esponja roja *Hymeniacidon sanguinea;* sería muy interesante saber si en aquellas especies que presentan dos o más morfos de color (como *Dendrodoris limbata,* con una forma "típica" pardo ocrácea y otra melánica) la segregación se ha debido a una diferencia alimentaria como en *Archidoris;* en la mayoría de los casos, sin embargo, se desconoce la especie alimento. Podrían multiplicarse los ejemplos de pares de especies, depredador y presa, cuya coloración es idéntica: *Rostanga pulchra/Ophlitaspongia pennata, Aldisa banyulensis/Hemimycale columella* (doridáceo/esponja), *Catriona aurantia/Tubularia indivisa* (eolidáceo/hidrario), etcétera. En los ejemplos indicados el color compartido es el rojo. En otros casos la homocromía se consigue a la vez por pigmentos propios y por los procedentes del alimento. Las manchas blancas del manto de *Bosellia mimetica* proceden de pigmentos propios, mientras que el color verde de este sacogloso se debe a la ingesta; sobre *Halimeda tuna* el molusco es perfectamente críptico, y las manchas blancas tienden a aumentar la semejanza porque imitan de alguna manera a los distintos organismos incrustantes que el alga presenta sobre sus frondes (poliquetos serpúlidos, hidrarios, briozoos y algas incrustantes). Un caso todavía más espectacular es el de *Serigea sublaevis,* de aguas canarias, que vive entre la esponja *Clathrina coriacea,* de la que se alimenta. El pigmento de la ingesta (amarillo) enmascara a este doridáceo sobre su sustrato habitual, pero la esponja es tubulosa, con filamentos anastomosados que dejan espacios huecos que aparecen al observador más oscuros que el resto de la colonia; este juego de sombras sería difícil de imitar si no fuera por la existencia en el dorso de *Serigea* de manchas ocráceas, de distribución irregular y producidas por pigmentos propios, que hacen que la semejanza sea perfecta. Esta misma esponja se presenta en otros dos morfos, de colores rojo y blanco; no se ha descubierto ninguna variedad de *S. sublaevis* con estos colores... pero ello quizá no sea más que una prueba de la eficacia del camuflaje.

Animales de otros muchos grupos de opistobranquios poseen pigmentos que se asemejan con mayor o menor perfec-

ción a la del sustrato, y en algunos casos coloraciones pretendidamente llamativas pueden no ser más que coloraciones crípticas sobre sustratos llamativos del mismo color; no obstante, no se conoce de manera suficiente el sustrato (o el alimento) de estas especies para aventurar una identidad cromática. En ocasiones, sin embargo, la semejanza es tan notoria que no sólo afecta al color, sino a la forma.

A esta homotipia u homomorfia se la ha llamado muchas veces, equivocadamente, mimetismo; veremos que el mimetismo consiste en otro tipo de imitación, todavía más complicada. Las semejanzas homotípicas más fácilmente conseguidas son las de eolidáceo-actinia (pues ya existe una cierta similitud morfológica entre los tentáculos de estos cnidarios y las papilas dorsales de los nudibranquios) y las de doridáceo-esponja, pero ciertos ejemplos son particularmente espectaculares. *Duvaucelia odhneri,* un dendronotáceo, presenta una elevada homomorfia con la gorgonia *Eunicella verrucosa,* y distintas especies de *Tritonia,* emparentadas con la anterior, con los alcionarios *(Alcyonium)* de que se alimentan: las ramificaciones del manto de los tritoníidos imitan perfectamente los pólipos abiertos de estos cnidarios, y la coloración contribuye a la cripsis. Otro tipo de semejanza participa a la vez de la homotipia y del comportamiento deimático, es decir, el destinado a asustar al posible depredador. Muchos eolidáceos, entre ellos *Dondice banyulensis, Hervia costai* y, sobre todo, *Spurilla neapolitana,* de la que ya se ha comentado su coloración semejante a la de las actinias y anémonas de mar que consume, adoptan una posición cuando se les molesta que hace difícil distinguirlos de un actiniario: el cuerpo queda totalmente oculto por una maraña de *cerata* que ondulan y amenazan con sus cnidosacos, de manera no distinta a como reaccionan a un contacto sospechoso la mayoría de actiniarios.

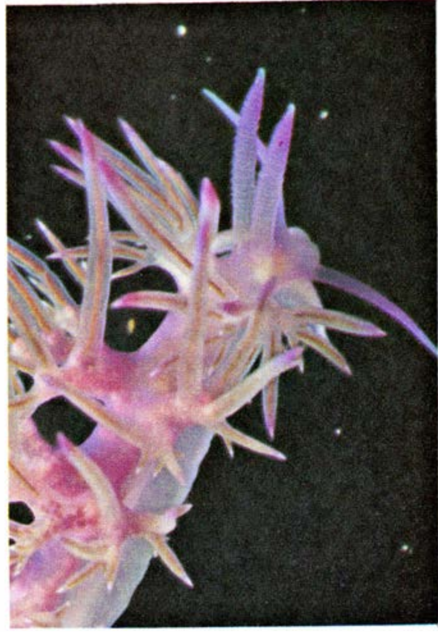
Este comportamiento posee asimismo otras ventajas: en el caso de que el depredador no resulte engañado por la semejanza (o intimidado por la amenaza combinada del aspecto y, en la mayoría de los casos, de la descarga de los cnidosacos) y muerda al eolidáceo, éste puede verse privado sólo de unos apéndices que son fácilmente regenerables, mientras que las partes más vulnerables del cuerpo se hallan escondidas bajo la batería defensiva que suponen los *cerata.* Otras semejanzas homotípicas son incluso más desconcertantes; tal es el caso de ciertos dendronotáceos pelágicos *(Scyllaea)* que imitan frondes de sargazos y el de algunos doridáceos tropicales *(Asteronotus)* parecidos a las hojas muertas que las aguas mueven sobre el fondo.

La cripsis sobre un sustrato determinado no precisa siempre de la imitación del color concreto de este sustrato; en ocasiones basta con que el animal críptico posea una coloración que "rompa" su silueta, es decir, que impida su identificación a los depredadores que reconocen a su presa porque coincide con cierta imagen mental que de ella tienen. La coloración disruptiva, típica por ejemplo de las hembras de muchas aves, se presenta asimismo en algunos doridáceos; en especial la he podido comprobar en *Peltodoris atromaculata.* Este gasterópodo de tamaño relativamente grande se alimenta exclusivamente de la esponja *Petrosia ficiformis,* sobre la que suele permanecer inmóvil por largos períodos de tiempo. Esta esponja es de color pardo rojizo debido a la presencia en sus tejidos de algas simbiontes, o completamente blanca cuando se halla en cuevas o desplomes oscuros donde los vegetales no pueden vivir; la coloración del doridáceo es blanco puro manchado irregularmente de pardo oscuro. Sobre su sustrato habitual el aspecto del conjunto doridáceo (o doridáceos)/esponja es el de una masa parda con manchas blancas, o el de una masa blanca con manchas pardas; estas "manchas" se confunden a veces con los restos de todo tipo y las incrustaciones orgánicas que pueden encontrarse sobre la esponja, pero en la mayoría de los casos bastan por sí solas para desdibujar o romper la silueta del animal y evitan una identificación rápida.

El camuflaje es perfecto, y se acentúa por el hecho de que el cuerpo del doridáceo no suele sobresalir del nivel de la superficie de la esponja, pues se instala en cavidades muy someras que excava en ella (a la manera de *Tylodina perversa,* aunque no tan profundas). Otros opistobranquios tienen coloraciones que serían "típicamente" disruptivas sobre determinados sustratos *(Dendrodoris limbata, Astureodoris rosi, Trapania maculata, Thecacera pennigera,* etcétera), pero casi siempre se desconocen éstos y, por tanto, no puede afirmarse que sus coloraciones abigarradas o moteadas se utilicen de igual modo que las de *P. atromaculata.*

El contrasombreado es un tipo de coloración muy general en los animales pelágicos, en especial en peces y cetáceos, por la que el dorso (o la parte del cuerpo que normalmente se halla en posición dorsal) es de color más oscuro que la parte ventral del animal; el resultado es que la superficie iluminada (dorsal) del cuerpo aparece con la misma tonalidad (generalmente azulada) y con idéntica intensidad luminosa que la parte ventral en sombra: el animal no destaca del entorno





DOS EOLIDACEOS *(Hervia costai, a la izquierda,* y *Flabellina affinis, a la derecha).* **El primero (que se desplaza con los** ***cerata*** **extendidos) se halla presto a la defensa, y ante cualquier perturbación dirigirá sus apéndices dorsales (en cuyo ápice se hallan los cnidosacos con acúmulos de los cnidocistos de sus presas, que son hidrarios del género** ***Eudendrium*****) hacia el enemigo potencial y expulsará las cápsulas urticantes. Mediterráneo (Blanes), fondos someros a profundos. El extremo anterior de** ***Flabellina affinis*** **muestra los extensos tentáculos orales y propodiales (lisos, cefálicos), los rinóforos sensoriales (lameliformes) y las papilas dorsales** ***(cerata),*** **en cuyo extremo se advierte la presencia de cnidosacos cargados de cnidos. La coloración contrastante del ápice de estos apéndices es advertidora, como la de** ***Hervia costai.*** **Mediterráneo, fondos de hidrarios de L'Estartit (Islas Medes).**





LA COLORACION DE LOS LOBULOS PARAPODIALES de este sacogloso *(Thuridilla hopei, a la izquierda)* **presenta un diseño claramente aposemático, mientras que el color general del cuerpo es verde azulado, homócromo con el alimento algal** *(Cystoseira)* **del gasterópodo. Mediterráneo, fondos de algas fotófilas de L'Ametlla de Mar.**

EL DISEÑO APOSEMATICO de *Glossodoris luteorosea (a la derecha)* **confiere a este doridáceo un bellísimo aspecto: sobre el dorso rojo carmín se distribuyen manchas irregulares amarillas orladas de rojo vivo, que hacen de este nudibranquio un animal extraordinariamente aparente. Cantábrico, fondos de esponjas de Fuenterrabía. Foto de J.D. Ros.**

acuático y puede pasar inadvertido a un depredador potencial. Los eolidáceos pelágicos (neustónicos, es decir, que viven pegados a la interfase agua-aire) están contrasombreados, en la posición anatómica normal en *Fiona pinnata* y en posición invertida en *Glaucus* y *Glaucilla,* que presentan el dorso claro y el vientre oscuro (azul, homócromo con el color de sus presas, los condróforos *Velella* y *Porpita* y el sifonóforo *Physalia).*

Se ha dicho que la descarga de las cápsulas urticantes de los glaucíidos puede ser fatal para el hombre, lo que no es sorprendente si se piensa que *Physalia* es un cnidario extraordinariamente bien dotado en lo que a sus armas defensivas u ofensivas se refiere. Ello nos introduce en otro sorprendente sistema defensivo propio de los nudibranquios: los eolidáceos que se alimentan de cnidarios y algunos dendronotáceos con el mismo régimen alimentario acumulan los cnidos o cápsulas urticantes de sus presas en unos cnidosacos de las papilas dorsales *(cerata).* Se ha llamado cleptocnidos a los nematocistos así obtenidos, pues se trata, efectivamente, de defensas "robadas" (cleptodefensas). En el opistobranquio las cápsulas pueden ser disparadas según un mecanismo que, en síntesis, es el siguiente: los cnidosacos se abren por acción mecánica y descargan los nematocistos, que abandonan las células que los nutrían (cnidóforos), son expulsados por el poro apical del *ceras,* se disparan al entrar en contacto con el agua de mar y actúan entonces con igual eficacia que cuando los expulsan los cnidarios.

Varios investigadores creyeron hasta finales del siglo pasado que las cápsulas urticantes de los eolidáceos eran propias y no adquiridas con la ingesta, pero aunque la demostración del verdadero origen de los cnidos no se hizo esperar, la resolución de la aparente "convergencia" defensiva entre depredador y presa no clarificó el todavía más sorprendente mecanismo por el que un arma eminentemente defensiva podía pasar, sin dispararse, al cuerpo de otro animal que la iba a utilizar, a su vez, con idénticos fines. Ya se ha comentado anteriormente cómo parece conseguirse esta transferencia o translocación, pero todavía no está muy clara la manera en que el opistobranquio utiliza para su beneficio este elemento disuasor tan eficaz para contrarrestar el ataque de muchos depredadores. La descarga de los cnidosacos se explica, alternativamente, por la acción mecánica o muscular (contracción de los músculos del *ceras*) del molusco, de una manera que podría calificarse, pues, de "voluntaria", o por la acción del depredador, que al presionar sobre las papilas provocaría el disparo de los cnidos.

Parece menos finalista y más de acuerdo con las observaciones efectuadas en el laboratorio considerar a los cnidosacos como unos depósitos de acumulación de estructuras demasiado peligrosas para que atraviesen todo el sistema digestivo del eolidáceo (algunas de ellas, como se ha dicho, son expulsadas con las heces, ya sea intactas o disparadas; pero parece que los cnidos que se acumulan son los más penetrantes o urticantes). La función defensiva habría aparecido más tarde en el curso de la evolución, con el perfeccionamiento de los cnidosacos y de su sistema de vaciado, sea éste el que fuere. Dos características adicionales de los *cerata* apoyan esta idea: por un lado, estos apéndices son autotomizables, es decir, el animal puede desprenderse de ellos a voluntad; por el otro, suelen ostentar coloraciones vistosas, en algunos casos homócromas con las de la presa debido a que se trata de acúmulos pigmentarios procedentes de la ingesta, pero casi siempre acompañadas o sustituidas por coloraciones producidas por pigmentos contrastantes propios, es decir, por coloraciones advertidoras.

La autotomía supone que no es necesario abrir los cnidosacos para liberar los cnidos: basta desprenderse de los *cerata* y dejar que el depredador los ingiera para que éstos actúen. (Incidentalmente, muchos otros opistobranquios, además de los eolidáceos, son capaces de autotomizar partes de su cuerpo –borde del manto en los doridáceos, apéndices dorsales y "cola" en los sacoglosos, etc.–. Estas partes del cuerpo que se "ofrecen" como tributo al depredador no sólo son regeneradas más tarde si el animal escapa ileso, sino que en ellas suelen ubicarse los sistemas defensivos propios de estos grupos: espículas, glándulas secretoras defensivas, etcétera.) Las coloraciones advertidoras son más generales en las especies más evolucionadas de eolidáceos, mientras que en las más primitivas, así como en las que no se alimentan de cnidarios, predominan las crípticas.

El conocimiento de la existencia de estructuras "robadas" en los opistobranquios (cnidos, pigmentos) parecía apoyar también la hipótesis de que las espículas de las esponjas utilizadas como alimento por algunos de ellos (en especial doridáceos) podían acumularse de igual modo. Ello no se ha comprobado, sin embargo, en ninguna especie, y las que presentan espículas ya se ha dicho que las poseen como estructura esquelética propia; por

otra parte, no se comprende cómo las espículas calcáreas de los dorídáceos podrían proceder de esponjas que, en su mayoría, las tienen silíceas.

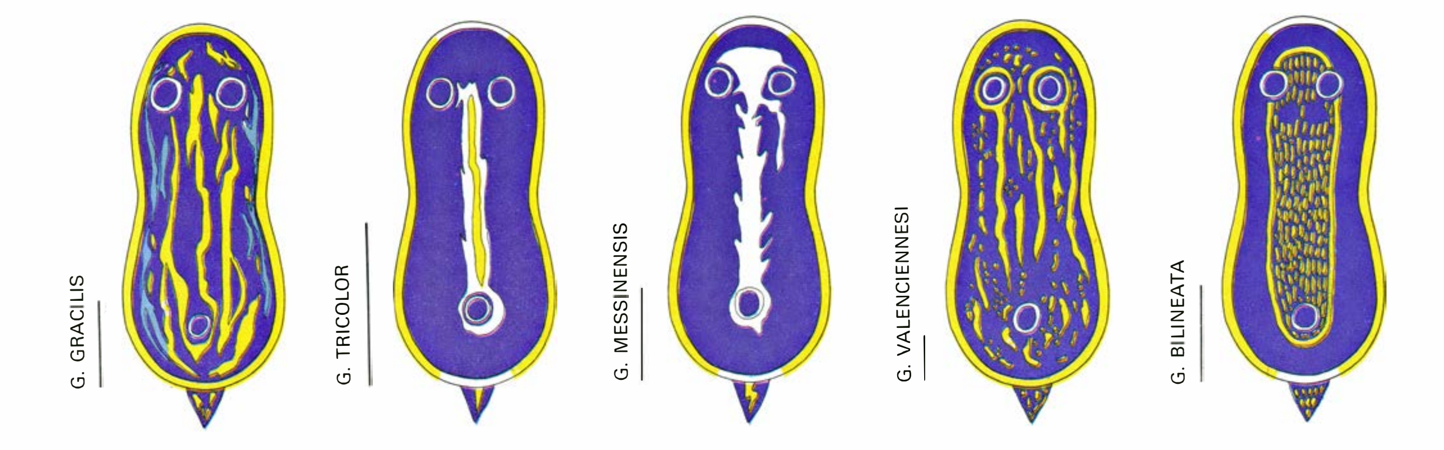
Insistamos en las coloraciones aposemáticas o advertidoras. Se ha dicho que las coloraciones vistosas de los opistobranquios pueden servir para atraer la atención del depredador hacia una parte del cuerpo defendida y fácilmente sacrificable, pero yo me inclino por otra utilización que parece más de acuerdo con el hecho de que muy pocos depredadores del mar comen a estos moluscos de coloraciones suntuosas. Como se ha dicho anteriormente, las coloraciones y los diseños de color aposemáticos obedecen a un fin único: hacer muy aparente al animal que los posee para que los posibles depredadores aprendan a reconocerlo y asocien la coloración vistosa con algún carácter tóxico, repugnante o desagradable en general; mediante aprendizaje, el depredador potencial acaba por "respetar" a los animales vistosos y éstos obtienen así una ventaja evidente de su llamativa librea. De hecho, parece que no sólo el color (y su exhibición mediante movimientos lentos) tiene valor advertidor, sino que también tendrían este carácter algunas sustancias repelentes que lo acompañan. Se comprende que si la defensa del animal se basa en un tóxico, que no tiene por qué tener un "sabor" desagradable, haga falta un repelente que sí lo tenga para que el depredador aprenda. El color (carácter superaposemático) avisaría del mal sabor (carácter aposemático), y éste del tóxico o de la defensa implicada.

Dejando aparte esta utilización, no probada para los opistobranquios, de las secreciones repelentes o desagradables, se ha especulado acerca de la posibilidad de que ciertos motivos aposemáticos no necesiten para ser eficaces de un aprendizaje previo por parte del depredador, pues despertarían en éste una repugnancia innata, similar a la que sienten algunos animales por las serpientes; esta interesante teoría no está tampoco probada para estos moluscos.

Si se piensa en el tipo de coloración que se necesita para que un animal resulte aparente y destaque de su entorno se advierte en seguida que no hay muchas posibilidades de elección; ya se ha comentado que un color vivo puede ser críptico sobre un sustrato igualmente coloreado (dorídáceo rojo sobre esponja roja, por ejemplo), y que ciertas combinaciones de color producen un modelo disruptivo, que también es críptico. Las posibilidades restantes se reducen prácticamente a una sola: la combinación de colores fuertemente contrastantes que destaquen, además, de la coloración del fondo; de ahí que en la mayoría de animales aposemáticos sean comunes combinaciones de negro y amarillo, en su defecto azul y rojo, por ejemplo. Insectos, reptiles y anfibios venenosos, peces defendidos por armas diversas y otros muchos animales terrestres y acuáticos presentan esta coloración contrastante, y los opistobranquios no son una excepción. No se descarta la posibilidad de que ciertas coloraciones o modelos de diseño, crípticos cuando el animal se halla sobre su sustrato habitual, se utilicen como advertidores cuando el entorno es distinto; se combinarían así dos posibilidades defensivas, lo que aumentaría las probabilidades de supervivencia de la especie correspondiente.

Muchos opistobranquios, en especial nudibranquios, ostentan coloraciones aposemáticas, y ello ha de entenderse que está relacionado con la eficacia de los diversos mecanismos de defensa que poseen estos animales y que ya se han comentado brevemente; debe señalarse, sin embargo, que para muchas especies poco estudiadas el carácter defensivo que justificaría la adopción de una librea aposemática se desconoce todavía. La coloración aposemática se restringe en algunas especies, como se ha dicho, a la parte del cuerpo que se halla implicada en la defensa: bordes del manto en *Doridium carnosum* (cefalaspídeo) y *Thuridilla hopei* (sacogloso), por ejemplo; papilas dorsales en los eolidáceos y *Fimbria fimbria* (dendronotáceo), apéndices dorsales de algunos dorídáceos; pero con frecuencia afecta a toda la parte visible del cuerpo (dorídáceos, principalmente, como *Glossodoris luteorosea*). Los ejemplos que podrían citarse para nuestras costas comprenden además algunos dendronotáceos y aplisiáceos (como *Aplysia dactylomela*, de color amarillento sobre el que contrastan círculos negros), y especies aisladas de los demás grupos.

Por la mismá razón de las reducidas posibilidades de elección entre pares de colores contrastantes se comprende que varias especies defendidas tiendan a adoptar un uniforme aposemático similar, a veces incluso idéntico. Se habla entonces de círculos aposemáticos, o de mimetismo de Müller, si bien en sentido estricto no se trate de ningún mimetismo. Aunque el origen de estos círculos de



POSIBLE CÍRCULO APOSEMÁTICO entre las especies mediterráneas "azules" del género *Glossodoris* (grupo *A* del texto). Los anillos de las figuras esquemáticas corresponden a rinóforos y branquias, que no se han dibujado para mayor claridad. La semejanza de los modelos de color es más notable si se consideran los juveniles de algunas especies (*G. gracilis* y *G. valenciennesi*), no representados aquí. El trazo significa 1 cm y todas las figuras se han hecho a partir de animales vivos o de fotografías de los mismos, y el color corresponde a los animales adultos.

advertencia puede ser la parquedad de diseños de colores contrastantes, su existencia implica una ventaja biológica adicional para los animales que los integran: éstos no sólo están defendidos, y presentan una coloración que así lo proclama, sino que además los ensayos de alimentación que los depredadores efectúan se reparten entre una población de presas mucho mayor, con lo que cada especie integrante del grupo aposemático aporta sólo una fracción del diezmo que correspondería a una sola especie advertidora. La selección natural, en base a esta ventaja evidente, ha tendido a uniformizar más aún las semejanzas aposemáticas entre las distintas especies que componen tales círculos.

En los glosodoríddos *(Glossodoris [=Chromodoris])* de las costas ibéricas parecen haberse establecido distintos círculos aposemáticos, integrados todos por especies congenéricas y extensibles quizás a otras afines no directamente emparentadas. Estos grupos, según el autor, serían hasta cuatro, caracterizados por la coloración general y las especies de doridáceos que se indican: *A)* Animales de color básico azul, con borde del manto y bandas longitudinales dorsales de color blanco, amarillo y/o amarillo anaranjado: *Glossodoris gracilis, G. tricolor, G. messinensis, G. valenciennesi* (morfo azul) y, en aguas marroquíes, *G. bilineata. B)* Animales de color básico rojo, rosado o grisáceo con manchas y puntos contrastantes amarillos: *G. luteorosea, G. krohni, G. luteopunctata* y *G. punctilucens* (estas dos últimas especies marroquí y canaria, respectivamente). *C)* Coloración uniforme o abigarrada blanca y/o amarilla: *G. valenciennesi* (morfo amarillo), *G. fontandraui, G. purpurea* y, por la gran similitud cromática, *Cadlina laevis* y *Diaphorodoris luteocincta* var. *alba. D)* Coloración de fondo blanca con manchas o papilas contrastantes rojas: *G. elegantula* y varias especies de otras familias de doridáceos; en la ilustración de las páginas 58 y 59 aparecen *Diaphorodoris papillata, Crimora papillata, Limacia clavigera, Polycera quadrilineata, Ancula gibbosa* y *Trapania maculata,* pero su inclusión se ha hecho sólo con carácter indicativo, para señalar que abundan las especies "moteadas" susceptibles de integrarse en un círculo como el indicado.

Naturalmente, estos círculos aposemáticos sólo pueden considerarse como tales si sus componentes comparten algún carácter defensivo y si conviven, por lo menos, en algunas zonas de su área de distribución. En cuanto al carácter defensivo de los *Glossodoris,* debe decirse que no está bien estudiado; se sabe que al menos las especies investigadas carecen de secreciones ácidas, pero deben poseerlas repelentes: no son aceptadas como alimento por peces y otros animales marinos, como se discute más adelante, y algunas de ellas poseen espículas. De los datos de distribución de los opistobranquios de nuestras costas, todavía incompletos, resulta evidente que estas especies coexisten, aunque algunas de ellas se hallan segregadas en cuanto al tipo de alimento y a algunos otros factores ecológicos (la luz, por ejemplo) en, por lo menos, algunas localidades bien estudiadas del Mediterráneo y del Cantábrico. Precisamente porque estas especies suelen encontrarse juntas y por parecerse mucho entre sí, algunos investigadores han resultado tan engañados como los depredadores a los que se supone va dirigida esta uniformidad aposemática: hasta hace poco tiempo se creía que *G. tricolor* y *G. messinensis* pertenecían a la misma especie, y los adultos de estas dos especies son muy difíciles de distinguir a simple vista de los juveniles de *G. gracilis* y *G. valenciennesi.* Los juveniles de *G. krohni* y *G. luteorosea* son asimismo aparentemente idénticos, y durante un siglo se ha creído que *G. elegantula* era una simple sinonimia de *Diaphorodoris papillata.*

Los representantes del círculo aposemático *D* parecen ser un claro ejemplo de mimetismo de Bates o, simplemente, mimetismo. Un animal (mimo o mimeta) mimetiza a otro (modelo) cuando adopta una coloración aposemática similar a la del modelo sin hallarse protegido como éste. Así comparte la inmunidad de que goza el modelo como animal aposemático y defendido. Es el típico caso de muchos dípteros inofensivos que imitan a las avispas, y el de mariposas comestibles que imitan a otras de sabor repugnante. En la ilustración de las páginas 58 y 59 se aprecia que distintas especies de doridáceos (aquí se han representado sólo algunas de las más comunes en aguas ibéricas) tienden a adoptar una coloración que corresponde a la de un eolidáceo defendido y, por tanto, con sus *cerata* coloreados. Pueden hallarse más ejemplos de nudibranquios con manchas o apéndices de colores vivos, distintos al rojo y anaranjado de los ilustrados.

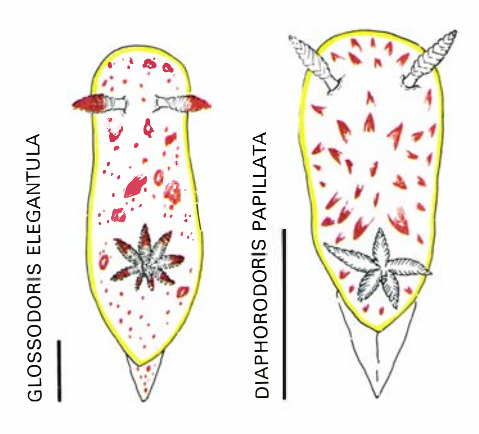
Estas agrupaciones aposemáticas de nudibranquios similares serían más numerosas si se ampliaran para las faunas de mares cálidos; existe ciertamente un elevado número de doridáceos "azules", "amarillos", "moteados", etc., en aguas del mar Rojo, del Caribe, de los océanos Atlántico, Pacífico e Indico tropicales, pero no parece que se hayan citado estos círculos aposemáticos que lógicamente, deberían hallarse mejor representados en las faunas tropicales, más ricas.

En las páginas que anteceden se ha ofrecido un repaso breve y forzosamente poco detallado de los métodos y estrategias de defensa en los opistobranquios; algunos de los casos citados están bien comprobados, otros parecen muy posibles y unos pocos son altamente especulativos. Precisamente para justificar la inclusión de aquellos sistemas defensivos que pueden parecer más subjetivos será bueno dedicar unos comentarios a las objeciones que generalmente se hacen a la existencia de tales mecanismos y, en especial, a los que emplean el color y las coloraciones adaptativas, es decir, el aposematismo y el mimetismo. La cripsis parece fuera de toda duda en los opistobranquios, como confirmaría cualquier recolector que haya intentado encontrar sacoglosos entre las algas de la costa.

La existencia de colores advertidores, reconocida en diversos grupos de animales, no siempre es aceptada en estos gasterópodos, y ello por tres tipos de razones fundamentales. Se dice, efectivamente, que:

1) En muchas especies que ostentan libreas de colores vivos no se ha probado la existencia de un carácter defensivo o desagradable para los depredadores; además, algunas especies de opistobranquios forman parte de la dieta de muchos animales del mar.

2) Muchas especies de opistobranquios viven ocultas, enterradas en el sedimento, bajo piedras o en otras situaciones donde tales coloraciones no son de ninguna utilidad.



EL MIMETISMO DE BATES se da cuando especies no protegidas por defensas propias imitan la coloración, el diseño de color y la for-

3) La visión acromática (que no distingue los colores) de los depredadores, la poca luz que llega a los fondos marinos y el hecho de que los opistobranquios encuentren su alimento por métodos olfativos y no visuales son distintos factores que se complementan para que la existencia de coloraciones adaptativas pueda ponerse en duda.

Con respecto al primer tipo de objeciones, puede decirse que muy pocas especies de opistobranquios han sido investigadas a fondo acerca de sus poderes defensivos y de la eficacia de los mismos, pero en la mayoría de ellas se ha encontrado la presencia de uno o más caracteres desagradables o "disuasores"; ello demuestra una vez más la versatilidad de todo el grupo, del que ya se ha dicho que, aunque de origen polifilético, ha evolucionado siguiendo líneas convergentes y ha producido varias veces, y de manera independiente, formas desnudas, con papilas, etcétera. Aboga asimismo por la existencia de caracteres defensivos en especies, afines o no, todavía no investigadas o que lo han sido parcialmente.

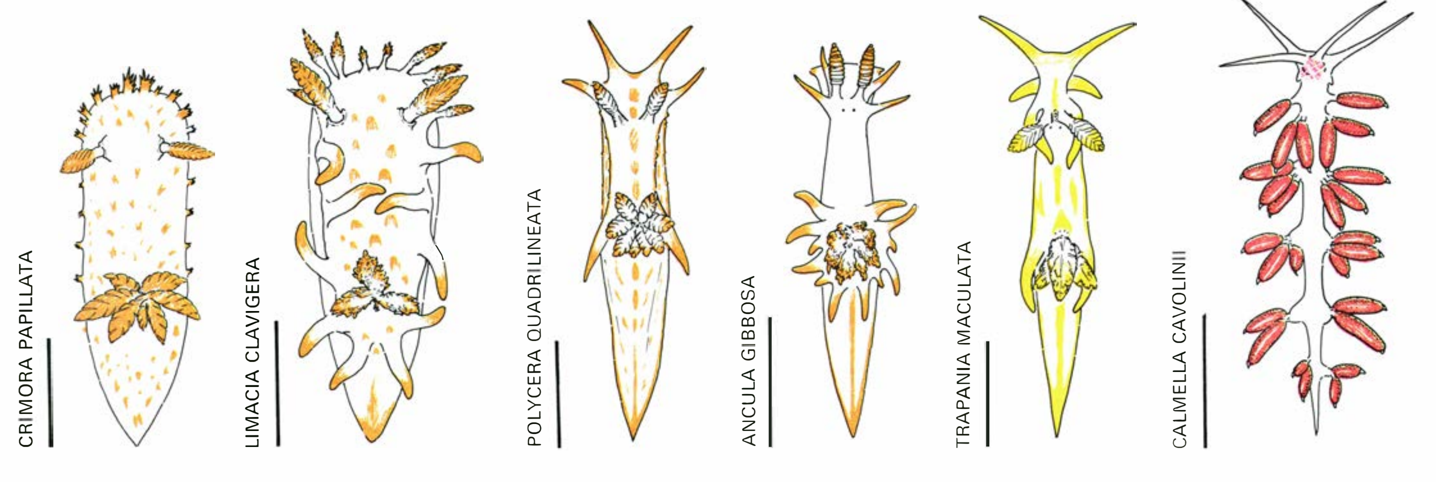
Las experiencias de laboratorio que suelen realizarse para comprobar la palatabilidad o comestibilidad de estos animales se basan casi siempre en el comportamiento de peces hambrientos frente a opistobranquios que se dejan caer libremente en el agua; la reacción de los peces y su aceptación o no de la presa indican si ésta se halla defendida. Se ha señalado, sin embargo, que el comportamiento de los peces no es un buen índice de la protección de un opistobranquio, puesto que los peces de acuario están habituados a recibir su comida en esta forma, mientras que para los peces en libertad una presa cayendo libremente a través de la columna de agua suele ser una visión insólita. El autor ha realizado una serie de experiencias, en acuario y en el mar, que intentan evitar estos artefactos experimentales que podrían desvirtuar por completo los resultados de una prueba de comestibilidad.

En una primera serie de pruebas, distintas especies de opistobranquios *(Umbraculum mediterraneum, Archidoris tuberculata, Glossodoris valenciennesi, Flabellina affinis)* se mantuvieron en grandes acuarios de exposición junto con peces (morenas, sargos, meros, serranos y tordos), crustáceos (bogavantes, langostas y cigarras de mar) y cefalópodos (pulpos) a lo largo de períodos diversos (hasta tres meses). Ninguno de los ejemplares de opistobranquios (que fueron depositados en el fondo, desde donde se desplazaron libremente, y no soltados desde arriba) fue aparentemente comido, mordido ni molestado mientras duró la experiencia, y sólo *U. mediterraneum* se ocultó bajo un saliente de piedra: los demás se movían por las paredes y el fondo del acuario. En otra serie de experiencias, esta vez en el mar, realizada en diferentes épocas, a profundidades y en lugares distintos de la costa catalana, se utilizaba un método de "atracción de peces" que conocen los buceadores de todo el mundo: si se rompen y desmenuzan erizos de mar, los fragmentos comestibles se esparcen y son pronto comidos por peces de distintas especies que acuden desde los alrededores, atraídos por el olor. Las pruebas se hacían soltando entre los restos de erizo distintos eolidáceos *(Flabellina affinis, Coryphella pedata, Hervia costai)*, doridáceos *(Glossodoris tricolor, G. gracilis, Peltodoris atromaculata)* y sacoglosos *(Thuridilla hopei)*. Tordos, doncellas, serranos, sargos y castañuelas, reunidos a veces en gran número, mordisqueaban en la nube de restos de erizo, pero en ninguna ocasión "confundieron" a los nudibranquios con aquéllos, ni en el agua ni sobre el fondo, a pesar de que el olor de los erizos debe ser penetrante (atrae a los peces desde gran distancia) y quizás impregna también a los nudibranquios.

Muchos tectibranquios y la mayoría de pterópodos, en general no defendidos y de colores apagados, son ampliamente consumidos por distintos depredadores marinos, y en el tubo digestivo de distintos peces se han encontrado en ocasiones algunas especies de nudibranquios; pero por lo general no son muchos los depredadores confirmados de los opistobranquios más defendidos, y repetidos ensayos en el laboratorio han dado resultados no distintos de los indicados más arriba, aunque mientras algunas especies no son tomadas nunca por los peces otras son probadas e inmediatamente rehusadas, en ocasiones varias veces.

En cuanto al segundo tipo de objeción, válida en términos generales, no tiene en cuenta, sin embargo, que las especies que suelen esconderse de algún modo no son precisamente las de colores más vivos (a excepción de las crípticas sobre sustratos llamativos, de las que ya se ha dicho que pueden hallarse también defendidas); es decir, las especies de opistobranquios con libreas aposemáticas comprobadas o presuntas suelen exhibirlas mediante hábitos que rara vez incluyen el ocultamiento.

La tercera objeción es quizá la de más peso, puesto que se basa en, por lo menos, tres argumentos que tenderían a invalidar



ma de otra (u otras) protegidas. Las especies representadas son doridáceos de cuerpo blanco translúcido con papilas o manchas rojas o anaranjadas, aspecto que les asemeja a un eolidáceo (como el representado a la derecha) con sus *cerata* pigmentados aposemáticamente (grupo *D* del texto). En éste se justifica que algunas coloraciones puedan ser anaranjadas o amarillas y no sólo rojas. El trazo representa 5 milímetros.

todo cuanto se ha dicho hasta aquí a propósito de los colores de los opistobranquios, y de los animales marinos en general. Veremos que no es así. En primer lugar, varias especies de peces de distintos grupos sistemáticos y algunos crustáceos y cefalópodos pueden ver los colores, y muestran determinadas preferencias cromáticas en su alimento o en su pareja sexual. Sin embargo, y ello enlaza con otra dificultad que se plantea a menudo, la visión de los colores no es en absoluto necesaria en el mar, donde a pocos metros de profundidad la luz del sol resulta absorbida y su espectro filtrado; a partir de una decena de metros el color predominante en el fondo es el azul o verdoso, que tiñe a organismos y sustrato con mayor o menor intensidad según sea su color propio. De ahí que sea razonable pensar, y ello se ha comprobado para algunas especies marinas, que los animales responden no tanto a los colores y a los motivos de color como a los modelos de contraste de un mismo color general.

Así, si se acepta que los círculos aposemáticos indicados anteriormente funcionan en cuevas y lugares medianamente oscuros o, en mar abierto, a profundidades de 20 m y más (muchos glosodorídidos y nudibranquios en general de nuestras costas se extienden hasta profundidades muy superiores, y el máximo de abundancia para todas las especies se sitúa entre los 15 y los 50 m), a estos niveles y en aguas claras mediterráneas el color blanco y el amarillo prácticamente no se distinguen entre sí y los dos aparecen como matices de blanco azulado; el anaranjado y el rojo siguen en la gradación que lleva a los verdes y azules, y todos ellos se ven como matices de un azul cada vez más oscuro, que a partir de los 40-50 m es general. El modelo aposemático no sería, pues, el que ofrecen los animales bien iluminados sino el que ostentan a estas profundidades (o, lo que es lo mismo, el que ven los depredadores, con visión del color o sin ella). Es decir, y volviendo a los cuatro grupos indicados anteriormente, las libreas advertidoras de estos animales tendrían el siguiente aspecto: *A)* Animales de fondo oscuro (azul), con bandas longitudinales claras contrastantes; *B)* animales de fondo oscuro (rojo, grisáceo), con manchas contrastantes claras; *C)* animales de fondo claro uniforme, contrastante sobre el sustrato oscuro; *D)* animales de fondo claro, con manchas contrastantes oscuras.

Quizás ahora se comprenda mejor por qué en el grupo *D*, por ejemplo, se han reunido animales con papilas o manchas de color rojo, anaranjado y amarillo; el efecto de contraste es el mismo. Esta idea viene apoyada además por otro hecho: algunos doridáceos que son rojos o rosados cuando adultos presentan una librea juvenil azulada *(G. krohni)*, mientras que otros *(G. luteorosea)* tienen morfos adultos rojos y azules. Se comprende también que la selección natural no haya impedido la amplia variación de detalle en los dibujos contrastantes (amarillos y blancos) del primer grupo, en el que es difícil encontrar dos animales de la misma especie exactamente iguales. En todos esos casos, la selección ha respetado todas aquellas variaciones menores que no perturbaban el modelo contrastante básico, que era el biológicamente eficaz.

Puede hacerse ahora un comentario sobre otra objeción, según la cual los propios opistobranquios no ven los colores, y, por tanto, no tendrían sentido en estos animales las coloraciones indicadas, en especial las crípticas. No se tiene cuenta que la cripsis, el aposematismo y demás, son fenómenos que han surgido como fruto de un largo proceso de selección natural, en el que los organismos implicados desempeñan un papel puramente pasivo y, por tanto, independiente de su capacidad visual. Es la visión de los depredadores la que cuenta, al seleccionar éstos los animales menos camuflados o de coloración aposemática menos perfecta.

Aunque puede aducirse que los fenómenos aposemáticos se dan en comunidades generalmente bien iluminadas y suelen estar dirigidos a animales de visión cromática, basta recordar que dos de las comunidades más ricas en ejemplos de coloración advertidora y mimetismo, la selva tropical y el arrecife de coral, son típicamente umbrías u oscuras. Teniendo en cuenta todo lo que acaba de señalarse, no resulta extraño el hecho (que tantas veces se ha esgrimido en contra de las coloraciones adaptativas, lo que no deja de resultar paradójico) de que las fotografías en blanco y negro resaltan más los fenómenos de cripsis o de aposematismo que las fotografías en color, aumentando el efecto de enmascaramiento o de advertencia, respectivamente. Y algunos opistobranquios aposemáticos están coloreados con diseños llamativos en negro y blanco.

De todo lo expuesto hasta aquí puede deducirse que los mecanismos de defensa en los opistobranquios, tanto por su naturaleza como por su eficacia, así como por el hecho de que dos o más de ellos pueden presentarse reunidos en una sola especie, pueden caracterizar a los componentes de este grupo al igual que otros caracteres biológicos (número de puestas por año, longevidad, etcétera) o ecológicos (tipo de alimento, de hábitat, etcétera). Es posible, por tanto, establecer una gradación, de menor a mayor eficacia, que de alguna manera refleje el "éxito" biológico de los distintos grupos considerados. Se tendría así: *1)* Animales aparentemente no protegidos y que no suelen ocultarse; son consumidos en cantidad por diversos depredadores: tecosomas, gimnosomas y muchos cefalaspídeos; *2)* animales protegidos (o no) que utilizan el ocultamiento en sus formas más simples; son normalmente especies excavadoras y de coloración inconspicua: tectibranquios bentónicos en general y algunos nudibranquios; *3)* animales protegidos o no que utilizan métodos de ocultamiento más refinados (cripsis, en todas sus variantes; anacoresis trófica): sobre todo sacoglosos y nudibranquios, pero también pleurobrancáceos y aplisiáceos. *4)* animales protegidos que no se ocultan (coloraciones y actitudes aposemáticas); la protección suele ser más elaborada y más eficaz que en los otros niveles: nudibranquios en general y algunos sacoglosos, pleurobrancáceos, aplisiáceos y cefalaspídeos. *5)* animales protegidos y no protegidos que comparten la coloración aposemática (mimetismo): algunos doridáceos; quizá representantes de otros grupos de nudibranquios.

Ciertamente, ninguna defensa es superior al mejor depredador, de manera que las excepciones existen y los opistobranquios adscritos a los dos primeros grupos, e incluso a los tres restantes, no se hallan a salvo del ciclo de la vida y de la muerte que supone, para la mayoría de animales, ser presa de otro en algún momento de su vida. Es indudable, sin embargo, que la evolución ha actuado para que este momento llegue en estos gasterópodos lo más tarde posible.

Se ha dicho que los opistobranquios son a los moluscos lo que las orquídeas a las angiospermas, las mariposas a los artrópodos y las aves a los vertebrados, significando con ello la belleza y variedad de sus magníficos colores. No es extraño que, como sucede para estos diversos grupos de seres vivos, el color tenga para los opistobranquios un significado que va más allá del "capricho" de la naturaleza. Muchos eones de evolución conjunta entre depredador y presa y entre presa y sustrato, en un ambiente estable y antiquísimo, han producido en ellos las semejanzas más perfectas y los diseños aposemáticos más estridentes y bellos de toda la serie animal.

La solidificación del cemento

¿Qué sucede cuando se añade agua al cemento para que una mezcla pastosa se transforme en una masa dura como la roca? Ciertas reacciones químicas producen un sistema de partículas entrecruzadas que unen la totalidad del material

D. D. Double y A. Hellawell

Lo que normalmente se llama cemento puede describirse como un material calcáreo, finamente pulverizado, que, al mezclarlo con agua, forma una pasta plástica que fragua y finalmente endurece hasta conseguir una consistencia semejante a la de la roca. Esto es lo que suele darse por supuesto, pero lo que no se percibe generalmente es que el desarrollo de la dureza es el resultado de las reacciones químicas entre los constituyentes del cemento y el agua añadida. El proceso de consolidación se debe a la formación de una rígida matriz engranada de productos de hidratación, que va sustituyendo poco a poco el agua existente entre los granos de cemento, hasta ligar finalmente toda la masa compuesta. Los procesos implicados en la hidratación son bastante complicados. El propósito de este artículo es describir algunos nuevos conocimientos en este campo, basados en nuestros propios estudios en la Universidad de Oxford y en los que otros profesionales han realizado en otras instituciones.

Para adquirir perspectiva, debe recordarse que el uso del cemento como elemento adhesivo entre ladrillos y piedras en la construcción de edificios y en trabajos de ingeniería no es, en absoluto, nuevo. Sus orígenes se remontan a Egipto y Grecia. El uso del cemento alcanzó un elevado nivel de pericia durante el período romano.

Ahora describiríamos con más precisión estos cementos primitivos como morteros de cal (mortero romano), es decir, mezcla de cal viva (óxido de calcio, CaO) y arena. Tanto el desarrollo de su solidez como su duración a largo plazo dependían no tanto de la reacción inicial en la que se apaga la cal –que se escribe $CaO + H_2O \rightarrow Ca(OH)_2 \downarrow$, para mostrar que, al añadir agua, precipita hidróxido de calcio y forma un coloide, sistema en el cual finas partículas del material están dispersas en el agua– como del proceso de envejecimiento subsiguiente por el que el hidróxido coloidal se transformaba lentamente en carbonato de calcio al reaccionar con el dióxido de carbono atmosférico $[Ca(OH)_2 + CO_2 \rightarrow \rightarrow CaCO_3]$.

La elevada calidad de los cementos romanos, que se hace evidente en el número y solidez de las estructuras romanas que aún siguen en pie, era debida, en gran medida, al descubrimiento suplementario de que la cal mezclada con material silíceo reactivo (en forma de cerámica molida o cenizas volcánicas) daban un cemento que desarrollaba una solidez y resistencia al agua superiores. El "cemento romano" así elaborado disfrutó de amplio prestigio y mantuvo su popularidad con pocas mejoras o evoluciones hasta finales del siglo XVIII.

Aunque haya discrepancias al respecto, la invención del cemento portland, tal como se conoce ahora, se atribuye generalmente a Joseph Aspdin, un constructor de Leeds que obtuvo una patente en 1824 para "un cemento de calidad superior que se asemeja a la piedra Portland". Su cemento se preparaba mediante la sinterización de proporciones fijas de piedra caliza o yeso (carbonato cálcico) con arcilla (aluminosilicatos) en un horno a temperatura muy alta. Este es el fundamento del método de fabricación que sigue aplicándose hoy en día. Durante el proceso de calcinación, las materias primas se combinan para producir clínkeres (pequeños grumos redondos, que posteriormente se muelen para obtener el polvo de cemento), que contienen silicatos de calcio reactivos. Los silicatos proporcionan al cemento su característica hidráulica, es decir, la propiedad de endurecerse mediante reacción con el agua. A diferencia de los primeros cementos basados en la cal, el tipo de cemento de Aspdin puede endurecer bajo el agua y también es resistente a la misma por mucho tiempo. Una de sus primeras utilizaciones en gran escala fue para la construcción del túnel del Támesis, desde el año 1825 a 1843, realizada por Marc Isambard Brunel y su hijo Kingdom.

El cemento portland ordinario sigue siendo, con mucha diferencia, el cemento más importante en términos de cantidad usada para propósitos de construcción. En su forma más extendida se le mezcla con material inerte de relleno (arena y materiales clasificados) para producir hormigón. Se han desarrollado distintas variantes para atender a exigencias específicas (como endurecimiento rápido, baja evolución de calor y resistencia a los sulfatos), pero todas ellas implican modificaciones pequeñas en la composición o el método de preparación.

Una notable excepción en este contexto es el cemento de alta alúmina, que también tiene propiedades hidráulicas aunque es bastante diferente en su composición química y característica de hidratación. Por esta razón proporciona un término de comparación interesante con el cemento portland. El cemento de alta alúmina, que fue desarrollado en Francia y apareció en el mercado comercial después de la primera guerra mundial, se fabrica mediante calcinación de piedra caliza con bauxita (óxido de aluminio hidratado).

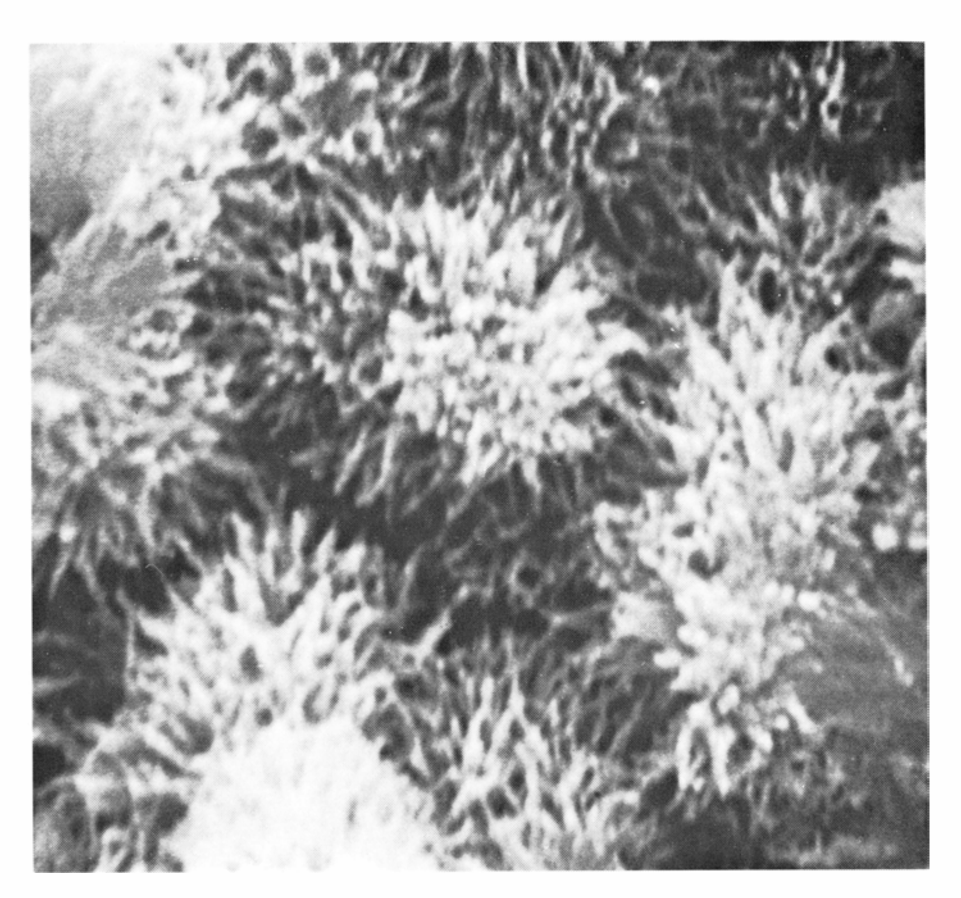
A diferencia del cemento portland, al que más o menos se llegó por ensayos y comprobaciones sucesivos, el cemento de alta alúmina fue diseñado específicamente para resistir la acción de los sulfatos. De hecho solucionó con éxito cierto número de problemas de construcción en aquellos medios donde no podía utilizarse el cemento portland debido a su vulnerabilidad frente al ataque corrosivo de los sulfatos. Otra ventaja del cemento de alta alúmina que pronto se apreció es que tiene una velocidad de endurecimiento muy rápida. Esta última cualidad lo ha convertido en un sustituto

interesante del cemento portland, cuyo proceso de endurecimiento es más lento, en aplicaciones tales como trabajos rápidos de reparación y en la fabricación de unidades premoldeadas, caso en el que una mayor velocidad de producción proporciona una ventaja económica que excede a su coste relativamente alto.

Sin embargo, la acumulación de experiencias con hormigones fabricados con cementos de alta alúmina ha puesto de manifiesto que su solidez a largo plazo puede dejar mucho que desear. Como resultado, el uso de este cemento para cualquier elemento que requiera capacidad de carga se ha convertido en una práctica muy discutida. En la mayoría de los países se evita ahora el uso de cementos de alta alúmina para propósitos estructurales, y algunos de ellos han impuesto prohibiciones o restricciones a este efecto. Es frecuente que estas acciones se hayan producido a raíz de incidentes en los cuales miembros estructurales y hasta edificios enteros hechos con cemento de alta alúmina se hayan desplomado, debido sobre todo al deterioro del hormigón.

En este artículo se describen las principales características de la hidratación y del desarrollo de la solidez del cemento portland y del cemento de alta alúmina, mediante un estudio comparativo. Ambos cementos se basan en un sistema ternario de óxidos de calcio, silicio y aluminio ($CaO–SiO_2–Al_2O_3$). Los cementos comerciales tienden a variar en composición dentro de un reducido margen y también contienen un cierto número de impurezas, sobre todo óxidos de hierro (menos del 10 por ciento en el cemento portland y hasta el 20 por ciento en el cemento de alta alúmina), y en grado mucho menor cal y óxidos asociados. Las impurezas, sin embargo, no afectan sustancialmente a las propiedades básicas de los cementos.

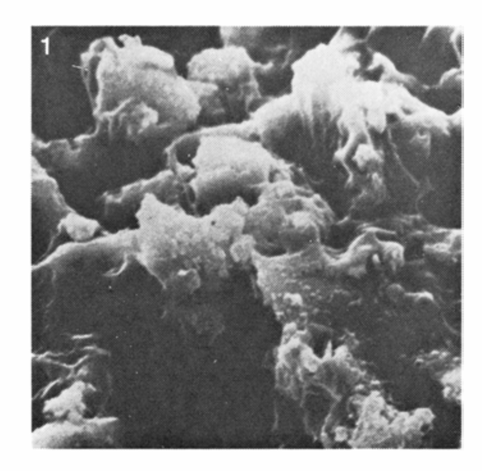
El cemento portland es básicamente una mezcla de silicato de calcio, sobre todo de los silicatos dicálcico y tricálcico, que pueden escribirse en forma abreviada C_2S y C_3S. (Emplearemos aquí una notación simplificada en la que los óxidos de calcio, silicio, aluminio y hierro están representados por C, S, A y F, respectivamente, y el agua por H. El silicato dicálcico, por ejemplo, que se escribiría normalmente $2CaO \cdot SiO_2$, en el sistema simplificado es C_2S.) También hay además proporciones más pequeñas de aluminato tricálcico (C_3A) y un ferroaluminato de calcio cuya fórmula aproximada es C_4AF. Las proporciones de peso típicas en un cemento portland ordinario

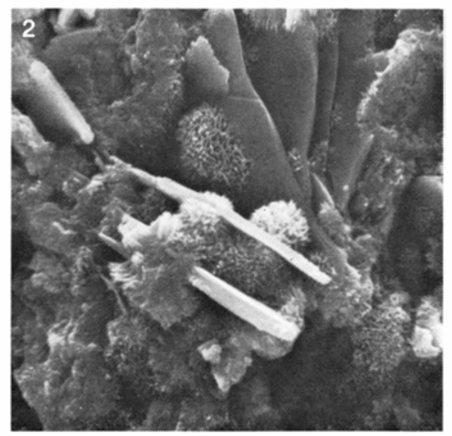


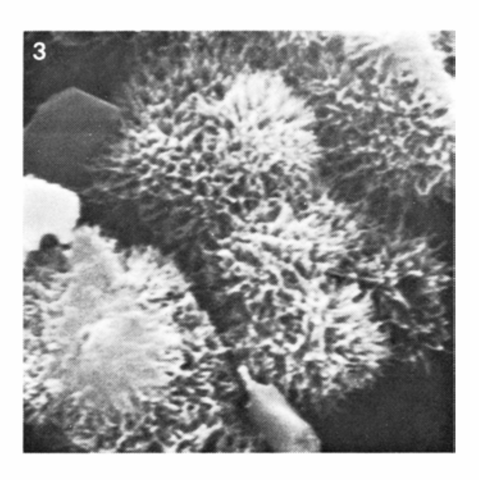
MICROESTRUCTURA DEL CEMENTO PORTLAND en las primeras etapas de la reacción con agua. Micrografía de barrido electrónico, hecha con un aumento de 6000 diámetros. Los granos individuales de cemento han empezado a desarrollar fibras del producto de reacción silicato-gel.



CEMENTO DE ALTA ALUMINA, diseñado para resistir el ataque de los sulfatos y que endurece más rápidamente que el cemento portland, fotografiado con el mismo aumento que la micrografía de la parte superior de la página. Los productos de reacción son aluminatos cristalinos con una forma entrecruzada acicular y laminar. La inestabilidad química de estos productos de reacción es la causa de un problema que presenta este cemento: su solidez no es predecible a largo plazo.



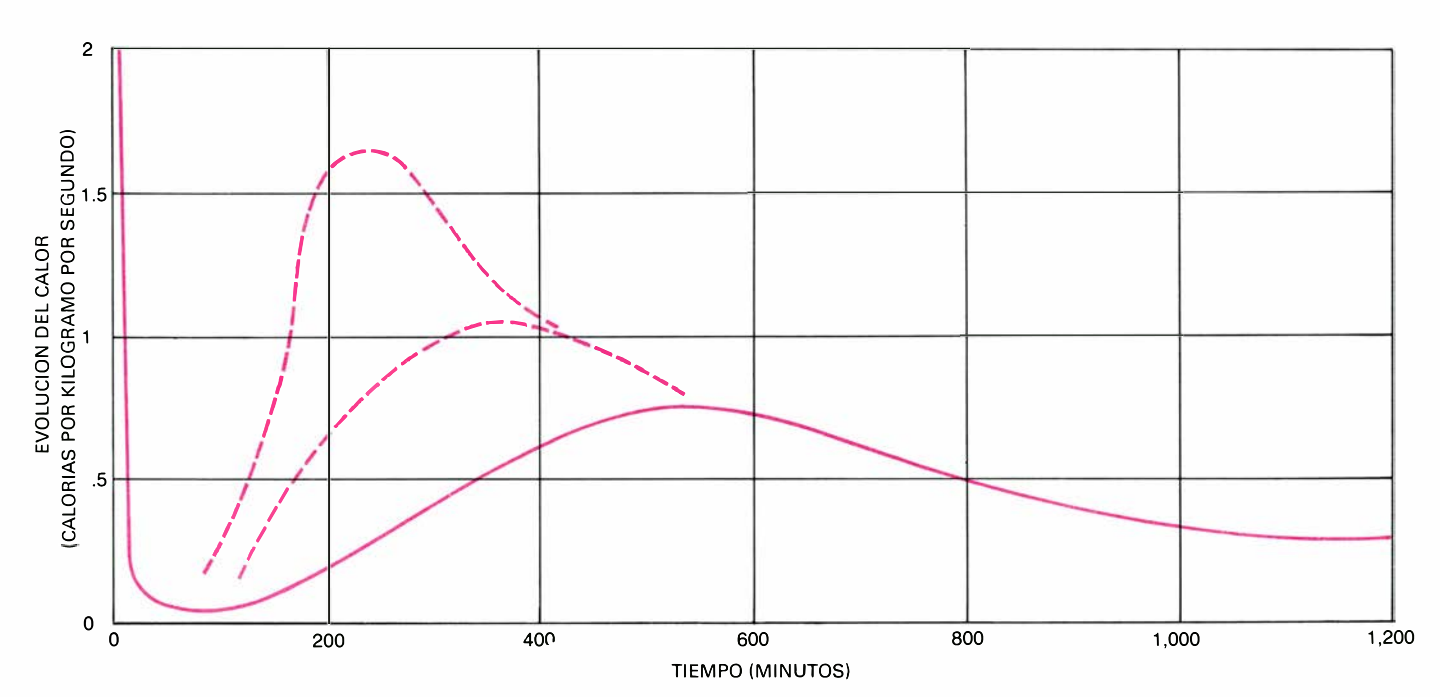




PASTAS DE CEMENTO PORTLAND aparecen en estas micrografías de barrido electrónico, realizadas en distintas etapas de hidratación, es decir, en diversos momentos después de mezclar el agua con el cemento. Al cabo de dos horas *(1)*, pueden verse las capas iniciales de gel alrededor de los granos de cemento. Después de un mes *(2)*, son evidentes las fibrillas, así como los grandes cristales laminares de hidróxido de calcio. Un detalle de la misma muestra *(3)* permite ver las fibras entrecruzadas. Estas micrografías han sido tomadas por los autores.

son 50 por ciento de C_3S, 25 por ciento de C_2S, 10 por ciento de C_3A, 10 por ciento de C_4AF y 5 por ciento de otros óxidos. El cemento de alta alúmina, como su nombre indica, contiene principalmente aluminato monocálcico ($CaO \cdot Al_2O_3$, escrito CA) y, relativamente, poco material silicato. También contiene otros varios componentes no tan bien especificados, pero CA es el principal ingrediente reactivo.

En ambos cementos, los principales constituyentes anhidros muestran propiedades hidráulicas: cuando se encuentran en forma finamente dividida, reaccionan con agua para producir una pasta que fragua y posteriormente endurece. Además, puesto que los cementos son diferentes desde el punto de vista químico, el modo y la velocidad de desarrollarse sus durezas son distintas. Ambos cementos pueden alcanzar, aproximadamente, la misma solidez última, pero el cemento de alta alúmina endurece inicialmente mucho más de prisa que el cemento portland. En la práctica, el cemento de alta alúmina puede desarrollar en un día la misma solidez que un cemento portland comparable lograría en un mes. Las ventajas del cemento de alta alúmina son, en este aspecto, obvias. El problema aparece al analizar su consistencia a largo plazo. Dependiendo de la temperatura ambiente y del medio, el cemento puede deteriorarse por un proceso conocido como conversión, de tal manera que la solidez presenta un dramático descenso repentino. Las condiciones que acentúan este comportamiento son una temperatura elevada y un contenido demasiado alto de agua en el cemento. Aunque estos factores también tenderían a reducir la consistencia del cemento portland a todas las edades, los efectos son mucho menos patentes y no se presenta un descenso repentino en la solidez comparable al desarrollado por el cemento de alta alúmina.



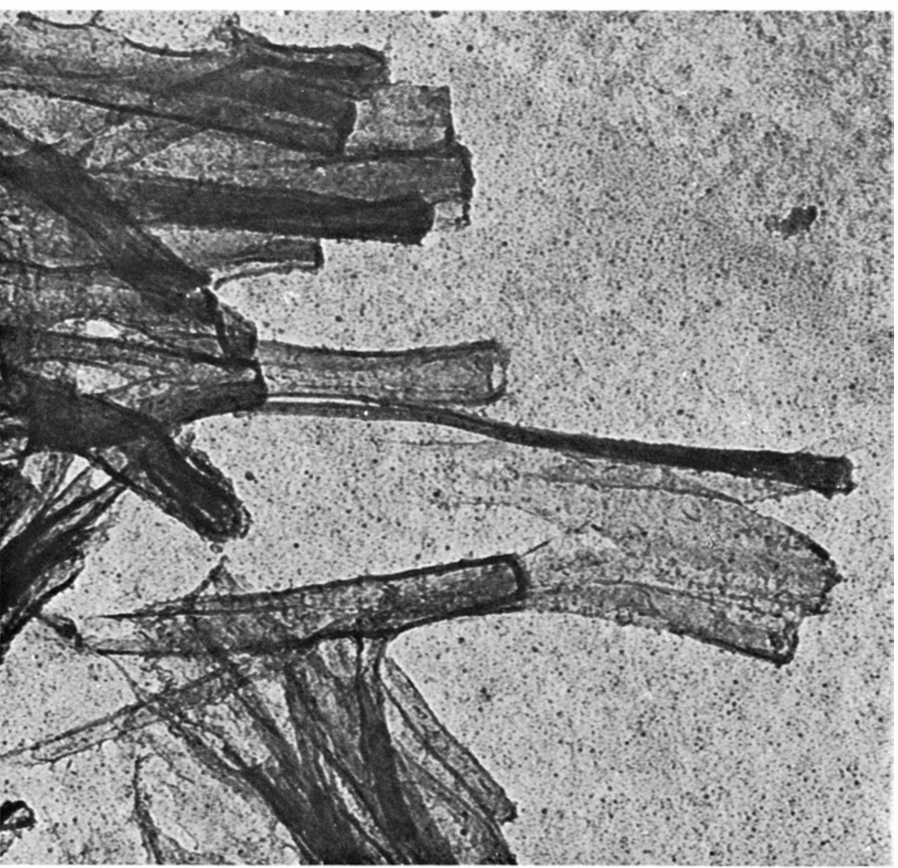
EVOLUCION DEL CALOR a medida que fragua y empieza a endurecer el cemento portland. El primer máximo agudo representa procesos rápidos de generación de calor, incluyendo la hidratación de la cal y la disolución de las impurezas del cemento. La zona plana que se encuentra a continuación representa el período "latente" o de "inducción", que probablemente es el resultado de una inhibición de la hidratación debida a los recubrimientos iniciales de gel, formados alrededor de los granos de cemento. Estos recubrimientos terminan perdiendo eficacia, porque la velocidad de la reacción alcanza pronto un segundo máximo. A una temperatura de hidratación de 20° Celsius, la velocidad de reacción está representada por la curva inferior. Las curvas discontinuas que aparecen por encima de ella muestran el efecto que se produce en la forma del segundo máximo al aumentar la temperatura de hidratación a 30 y 40 grados Celsius, respectivamente, según puede apreciarse.

Tomando como composición del cemento portland la dada anteriormente, procede considerar lo que conocemos acerca de los productos de hidratación y sus proporciones en la pasta endurecida. Tal y como podía esperarse, la reacción de endurecimiento en una pasta agua-cemento está asociada principalmente con la hidratación de los dos componentes silicatos, C_3S y C_2S. En el clínker anhidro ocupan alrededor del 75 por ciento del peso total y proporcionan la mayor parte del material hidratado. Las velocidades de reacción del C_3S y C_2S difieren apreciablemente, pero el producto final se considera el mismo en ambos casos: un gel coloidal de hidrato-silicato de calcio de composición y estructura bastante indefinidas. En un cemento endurecido, este gel, en una forma amorfa y finamente dividida, ocupa alrededor del 70 por ciento en volumen del material hidratado y, en consecuencia, constituye el principal agente de unión entre los granos residuales de cemento que no han reaccionado y los otros productos cristalinos de hidratación. La mayor parte de estos productos cristalinos consiste en hidróxido de calcio (portlandita), y el resto se compone de varios hidratos de aluminio complejos y otras sustancias menores.

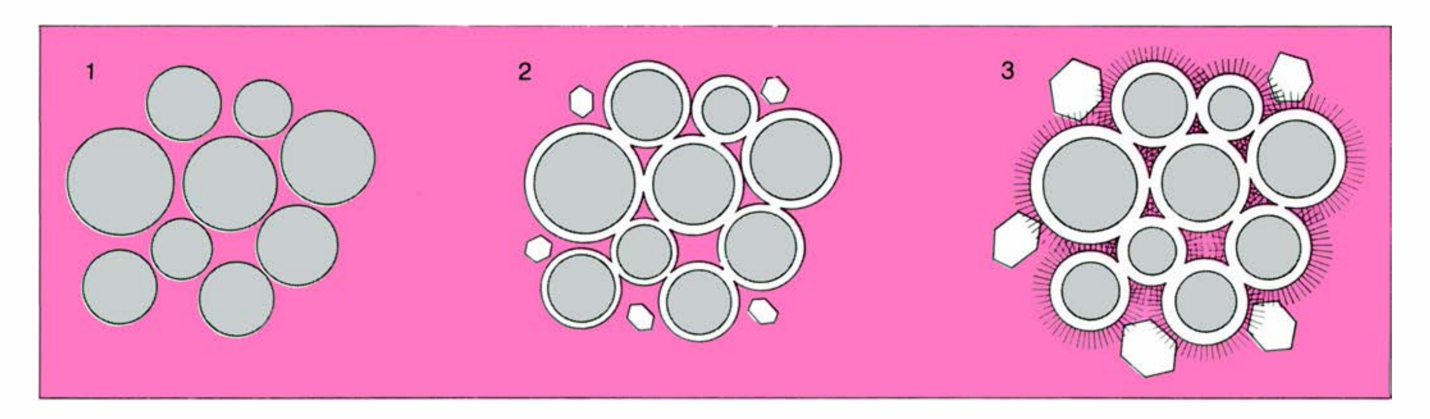
De los dos silicatos, C_3S está asociado primordialmente con la reacción de endurecimiento que acontece durante las primeras etapas de hidratación, dentro de los primeros días y semanas. El C_2S, que reacciona más lentamente, es probablemente el responsable de la solidez a largo plazo y de su continuo desarrollo, que puede extenderse a períodos de meses o incluso años. La contribución de los componentes aluminatos a la consistencia es incierta y relativamente pequeña.

En una pasta de cemento bajo condiciones normales, la reacción de hidratación no llega generalmente a ser completa. Incluso después de algunos años, una microestructura típica presentará granos residuales de cemento, que no han reaccionado, incrustados en una matriz compuesta principalmente de gel de hidrato-silicato de calcio. La causa de este fenómeno radica en que, durante la hidratación, los productos de reacción forman recubrimientos alrededor de los granos de cemento, recubrimientos que inhiben progresivamente el acceso de agua al material anhidro. Solamente puede alcanzarse la hidratación completa, en un lapso razonable de tiempo, mediante la ruptura continua de los recubrimientos para que queden expuestas nuevas superficies reactivas, como podría hacerse, por ejemplo, mediante una agitación continua de la pasta en el molino de bolas.





CEMENTO PORTLAND HIDRATADO, observado en su microestructura mediante el microscopio electrónico de transmisión. La micrografía superior presenta una muestra húmeda de cemento después de unos dos días de hidratación; se puede observar un crecimiento fibrilar bien desarrollado de gel silicato alrededor de los granos de cemento. El aumento es de 20.000 diámetros. En la micrografía inferior (aumentada 70.000 diámetros) aparece una muestra seca de cemento hidratado; ésta, como puede observarse, manifiesta con absoluta nitidez la naturaleza tubular de las fibras.



ETAPAS DE HIDRATACION del cemento portland, representadas esquemáticamente. El proceso empieza *(1)* con la adición de agua, que rodea los granos de cemento. Pronto aparece el recubrimiento de gel alrededor de los granos *(2)* y surgen los cristales angulares de hidróxido de calcio, que se desarrollan como un subproducto de la hidratación de los silicatos de la mezcla. Finalmente, se desarrollan las fibrillas tubulares *(3)*. El entrecruzamiento de estas fibrillas a medida que prosigue la hidratación termina amasando el cemento y los otros componentes de la mezcla.

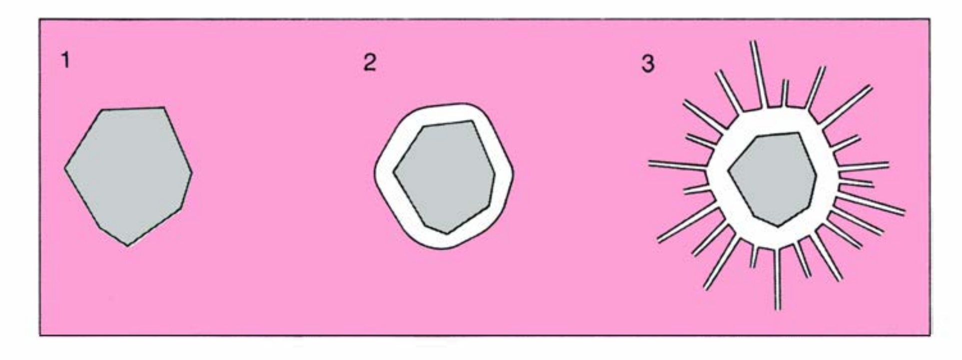
La presencia de material clínker sin reaccionar puede tener algunos efectos interesantes. Por ejemplo, si se muele el cemento y se mezcla otra vez con agua, fraguará por segunda vez y desarrollará un grado de solidez que, aunque es inferior al del primer fraguado, puede, sin embargo, ser importante. Es posible incluso que, si se desarrollan finas microgrietas en una estructura de cemento, lleguen a cerrarse por sí solas mediante una hidratación, renovada.

La reacción entre el cemento y el agua es exotérmica; bajo condiciones típicas, la temperatura de una pasta puede elevarse decenas de grados. Este fenómeno puede presentar un problema al instalar estructuras de hormigón de sección ancha. A veces se hace necesario enfriar el material durante el período de fraguado y endurecimiento, para evitar grietas por expansión y contracción térmicas. El calor emitido está directamente relacionado con el grado de hidratación de los distintos constituyentes del cemento. Debido a la naturaleza física de la reacción, la velocidad de evolución del calor disminuye exponencialmente para largos intervalos de tiempo. Sin embargo, bajo condiciones normales, alrededor del 50 por ciento del calor total disponible (100 a 150 calorías por gramo) se libera dentro de los primeros tres días de mezcla y cerca de un 30 por ciento en el primer día.

Si la evolución del calor se representa gráficamente en forma de curva, ofrece dos máximos. Un máximo es agudo, lo cual refleja procesos muy exotérmicos; se alcanza dentro de los diez minutos de hacerse la mezcla y se debe principalmente a la hidratación de la cal y a la disolución de otras impurezas que están siempre presentes, en pequeñas cantidades, en el cemento portland comercial. Después de este primer máximo, la velocidad de reacción decrece rápidamente hacia un valor bajo, donde permanece durante una o dos horas. Este es el período "latente" o de "inducción". Se piensa que se produce por la inhibición de la hidratación a medida que los granos de cemento son recubiertos por los productos gelatinosos iniciales. Este modelo se ve confirmado por observaciones realizadas con el microscopio electrónico.

Más tarde, las capas parecen perder eficacia, porque la velocidad de reacción empieza a incrementarse de nuevo y alcanza un segundo máximo (mucho más bajo) a las ocho o diez horas. Después decrece gradualmente en un período mucho más largo. La forma de este máximo está determinada fundamentalmente por la hidratación del constituyente mayoritario del cemento, el silicato tricálcico. La curva tiene una forma característica de muchas transformaciones de fase producidas por un proceso de nucleación y crecimiento. Tal y como podría esperarse, el grado de la reacción aumenta con la temperatura y está directamente relacionado con el área superficial del polvo del cemento, es decir, con la mayor o menor finura con que se le ha molido. Las variaciones en la proporción de agua a cemento comprendidas entre el 30 y el 50 por ciento (que son típicas en la práctica) no alteran significativamente la pauta de hidratación.

Aunque las curvas de evolución de calor informan sobre el grado y la velocidad de hidratación, dicen poco acerca de la estructura del material y no están necesariamente relacionadas, de un modo simple, con el desarrollo de la solidez. Si bien es cierto que, bajo condiciones normales, el período de fraguado ocupa aproximadamente el intervalo de tiempo que media entre el período latente y el segundo máximo, y que la reacción de endurecimiento aparece después, estos procesos de fraguado y endurecimiento dependen claramente del desarrollo detallado de la microestructura de la pasta. Debido al reducido tamaño de los hidra-



LOS "JARDINES" DE SILICE proporcionan una analogía al crecimiento de fibrillas tubulares en el cemento portland. Se forma un jardín de sílice cuando las sales solubles en agua de ciertos metales se introducen en una solución de silicato de sodio. En este ejemplo, se parte de un cristal de nitrato de cobalto. El cristal de sal empieza a disolverse *(1)* y a su alrededor se forma un caparazón de silicato insoluble *(2)*. Puesto que la membrana es permeable al agua, el cristal continúa disolviéndose. Finalmente, la presión osmótica rompe la membrana *(3)*, causando el crecimiento del silicato precipitado en forma de finos tubos horadados, similares a los que se encuentran en el cemento portland hidratado, aunque en un jardín de sílice las fibrillas son mucho más grandes.

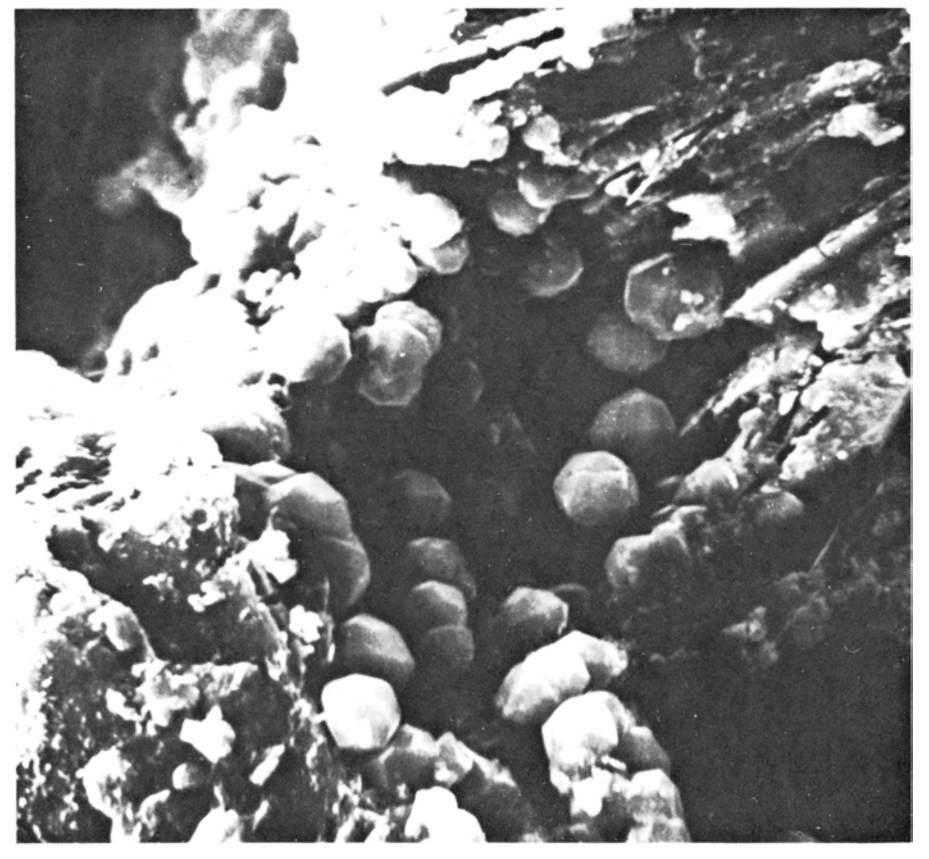
tos implicados, el estudio de la microestructura con el microscopio electrónico resulta muy ventajoso.

Las micrografías muestran que, en el plazo de dos horas, los granos de cemento, originalmente angulosos y cristalinos, van siendo envueltos por un recubrimiento gelatinoso de material hidrato. Los granos están unidos por sus puntos de contacto mediante estas capas. La red de ligaduras relativamente débiles así formada produce la pérdida de plasticidad de la pasta por el efecto de fraguado. La estructura puede romperse fácilmente, pero volverá a fraguar otra vez con bastante rapidez a medida que los granos de cemento se recoagulen.

El desarrollo irreversible de la solidez, durante el período de endurecimiento, que empieza después de tres a cinco horas, está asociado con un neto cambio en la microestructura del cemento. Los recubrimientos iniciales de gel desarrollan finas protuberancias superficiales. Durante un cierto tiempo van creciendo hasta llegar a formar unas fibrillas delgadas, densamente comprimidas, que irradian de cada grano de cemento a semejanza de las púas de un erizo. Este desarrollo fibrilar es el gel hidrato-silicato de calcio, producto de la reacción entre los silicatos del cemento y el agua. Al mismo tiempo, el hidróxido de calcio (portlandita) precipita como un residuo y forma los grandes cristales angulares que son tan prominentes en la microestructura. A medida que se produce la hidratación, las fibrillas de gel interpenetran gradualmente en la región de los granos de cemento adyacentes. La malla entretejida, así creada, se consolida posteriormente formando una matriz rígida que une toda la masa compuesta: los granos residuales de cemento que no han reaccionado, los productos cristalinos de hidratación y, en un hormigón, la arena y los materiales agregados.

Aunque el crecimiento secundario de las fibrillas de gel hidrato-silicato de calcio es de una clara importancia en el desarrollo de la solidez de la pasta agua-cemento, su morfología detallada y su mecanismo de crecimiento han permanecido oscuros hasta ahora. Estudios recientes realizados por nuestro grupo de investigación en Oxford, haciendo uso del microscopio electrónico de transmisión, han proporcionado una nueva comprensión de este problema. Estas imágenes muestran que las fibrillas no se asemejan a formas de caras regulares, tal y como se podía esperar de un crecimiento de cristales sólidos a partir de soluciones acuosas, y que las fibrillas individuales no son sólidas, sino que aparecen





DESARROLLO DE LA MICROESTRUCTURA de la pasta hidratada de cemento de alta alúmina, observado en el microscopio electrónico de barrido. La pasta fue elaborada con una alta proporción de agua a cemento (60 por ciento) y almacenada a alta temperatura (40º C.) para estudiar el proceso de conversión, que a menudo produce la quiebra de los miembros de carga hechos con cemento de alta alúmina. La micrografía superior muestra la pasta despues de un día; los cristales aciculares de los productos iniciales de hidratación metaestables son evidentes. La micrografía inferior, a los 14 días, muestra los poliedros facetados del producto de conversión, estable y débil.

como finos tubos horadados. Esta información presenta un aspecto completamente nuevo del problema.

El concepto de una morfología tubular al referirse a la precipitación de sólidos a partir de soluciones acuosas es tal vez bastante insólito. Resulta, sin embargo, que estas formas de crecimiento son más comunes de lo que podría esperarse. Un ejemplo familiar de crecimiento tubular lo encontramos en los "jardines" de sílice, que se desarrollan al introducir sales solubles de metales, que forman por sí mismas silicatos insolubles, en una solución acuosa diluida de silicato de sodio. El modelo de crecimiento de un jardín de sílice proporciona una analogía visual sugestiva con lo que sucede durante la hidratación del cemento portland. Primero se precipita alrededor del cristal de sal un recubrimiento gelatinoso de hidrato-silicato. Después de un período latente, el gel se abre intermitentemente para producir finas formas de crecimiento tubulares con una apariencia de irradiación fibrilar. A pesar de ciertas diferencias obvias entre la hidratación del cemento portland y el crecimiento de un jardín de sílice (en la escala de crecimiento y la naturaleza de los constituyentes químicos), los dos procesos son tan parecidos que sugieren que en ambos casos opera un mecanismo de crecimiento similar.

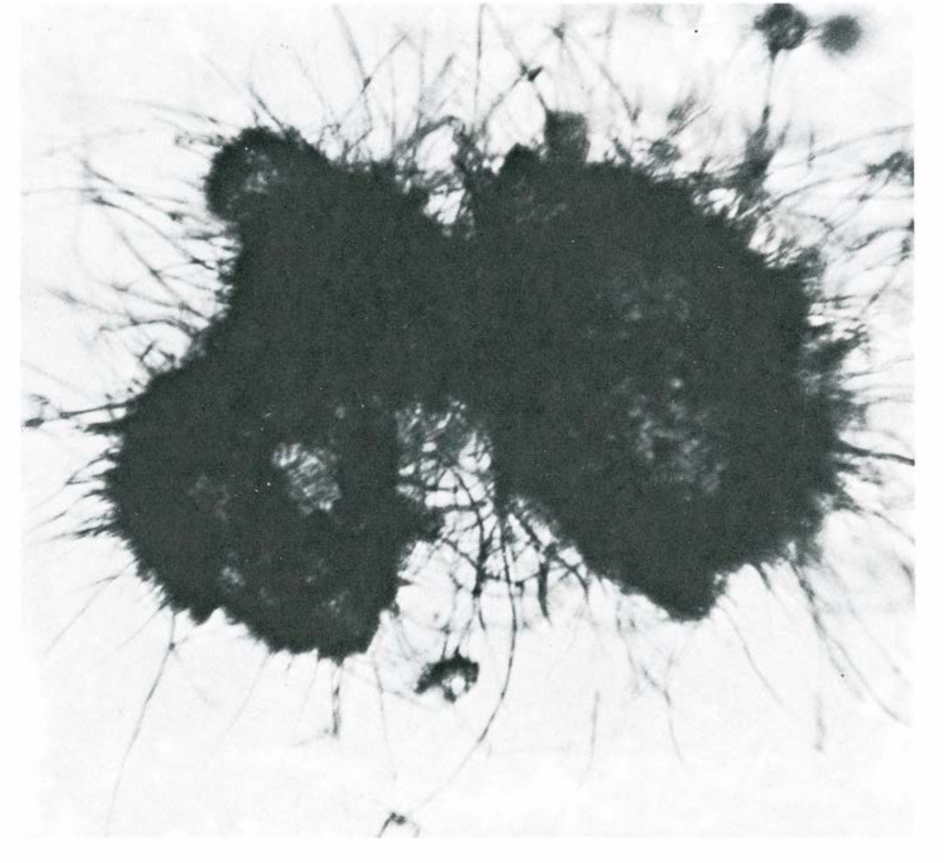
Se ha demostrado que el mecanismo de crecimiento del jardín de sílice depende de la presión osmótica. A medida que los cristales de sal se ponen en contacto con la solución acuosa de silicato, empiezan a disolverse y simultáneamente precipita un envolvente de hidrato-silicato. Este recubrimiento es coloidal y actúa como una membrana semipermeable. Debido a la diferencia de concentración entre las soluciones existentes a ambos lados de la membrana (una solución concentrada de la sal del metal en el interior y una solución diluida de silicato en el exterior), el agua se difunde preferentemente hacia el interior del envolvente, provocando una disolución continuada del cristal de sal. La presión interna aumenta lentamente y, por último, el recubrimiento explota, lanzando chorros de solución de sal dentro de la solución de silicato. Hay una reacción espontánea entre las dos, con precipitación del material silicato insoluble en forma de pared tubular continuamente creciente. La membrana inicial actúa como una bomba osmótica; la solución de sal fluye a través del tubo y precipita en el extremo mediante un proceso más o menos regular.

Hasta ahora hay pocos datos cuantitativos que orienten sobre este modo de crecimiento, pero es probable que sea relativamente común en aquellos casos en que el producto de la reacción inicial entre dos soluciones sea una membrana semipermeable continua. El recubrimiento inicial y los tubos que crecen del mismo son ambos amorfos y gelatinosos, aunque cuando se les extrae de la solución resultan ser bastante duros y fuertes. En cualquier sistema dado, la velocidad de crecimiento es proporcional al diámetro medido del tubo; los tubos más pequeños son el resultado de un crecimiento más lento. Este dato está cualitativamente en la dirección correcta para sostener la analogía entre los jardines de sílice y los productos de hidratación del cemento portland. Una amplia variedad de sales de ion metálico presentan este tipo de crecimiento. Es tranquilizador que las sales de calcio estén entre ellas, puesto que el gel de hidrato-silicato de calcio es el principal producto de hidratación del cemento portland.

Usando como modelo el jardín de sílice es posible explicar los aspectos esenciales de la hidratación del cemento portland y, en particular, el crecimiento del gel fibrilar de hidrato-silicato de calcio, que es un factor tan importante en el desarrollo de la solidez. La analogía implica que es la ósmosis la que proporciona la fuerza motriz para llevar a cabo la hidratación del cemento portland.

Los productos de hidratación del cemento de alta alúmina son formas cristalinas más convencionales. Tal como hemos indicado, el principal constituyente anhidro del cemento de alta alúmina es el aluminato monocálcico ($CaO \cdot Al_2O_3$, escrito CA), y el desarrollo de la solidez del cemento se deriva primordialmente de los productos de hidratación de este componente. Bajo condiciones normales, a temperatura ambiente, el principal producto es la fase decahidrato CAH_{10}, que está generalmente asociada con pequeñas cantidades de octahidrato C_2AH_8 con gel de alúmina, que con el tiempo se transforma en yesita cristalina, AH_3. Las cantidades relativas de CAH_{10} y C_2AH_8 presentes en la pasta de cemento dependen de la temperatura de hidratación y de la composición del propio cemento. Estos hidratos cristalinos proporcionan la matriz que une toda la pasta. Su pronta producción conduce al rápido incremento de la solidez del material.

La conversión acontece debido a que los hidratos iniciales CAH_{10} y C_2AH_8 son metaestables y tienden a cambiar espon-



UNA ESTRUCTURA MEJORADA que podría buscarse para el cemento portland viene sugerida por la fina red de fibrillas tubulares densamente entretejida visible en esta micrografía. El cristal, aumentado en 32 diámetros es de sulfato ferroso, sumergido antes en una solución de silicato de sodio. Un entrecruzamiento de fibras más denso mejoraría la resistencia a la tracción del portland.

táneamente al hidrato C_3AH_6, que es más estable, produciendo agua y yesita como residuos. El proceso va acompañado por un cambio en la estructura del cristal de los hidratos (CAH_{10} y C_2AH_8 tienen estructuras cristalinas hexagonales bien definidas y el C_3AH_6 es cúbico) y, debido a las diferencias en sus densidades, por un decrecimiento sustancial en el volumen de los sólidos. La modificación viene a ser de un 50 por ciento para el CAH_{10} y de un 30 por ciento para el C_2AH_8.

Estos cambios son la base del problema que se presenta en la solidez a largo plazo del cemento de alta alúmina. Conducen a una pérdida de cohesión en la microestructura y a un incremento de la porosidad de la pasta de cemento endurecida. A temperatura ambiente y a temperaturas más bajas, la velocidad de cambio es lenta, sobre todo en condiciones secas, y los efectos pueden no ser evidentes durante años. A temperaturas más altas, y en un medio ambiente húmedo, los cambios pueden acontecer en unas pocas semanas o incluso días. Al fin, el resultado es una marcada reducción de la solidez y una menor resistencia al ataque corrosivo. El cemento de alta alúmina dura más tiempo cuando se mezcla bastante seco y en un medio ambiente frío. Las combinaciones contrarias son desastrosas para la estabilidad del material a largo plazo.

Pueden seguirse las reacciones examinando, con el microscopio electrónico de barrido, muestras de pasta de cemento en varios estados de hidratación. Nuestras muestras, que estaban hechas con una alta proporción de agua a cemento, se envejecían artificialmente almacenándolas en un medio ambiente húmedo a una temperatura considerablemente alta (40° Celsius). Al principio, el cemento se recubre con una suave capa gelatinosa, no muy diferente en apariencia de la que se forma en el cemento portland. Después de un corto período de incubación, las capas se agrietan con unos surcos que rápidamente se transforman en cristales con caras de contorno hexagonal o acicular (como una aguja). Estos cristales son los primeros productos de hidratación; su complejo entramado da al cemento su rápida solidez.

Bajo las condiciones de envejecimiento simuladas, la muestra sufre una rápida transformación y en 14 días presenta la forma más estable de un producto de hidratación: el hexahidrato de forma casi esférica con poliedros facetados. Puesto que los cristales facetados son característicos del crecimiento a partir de una solución, la conclusión a que se llega es que el cambio al hidrato estable debe realizarse en esta forma. En otras palabras, el cambio implica una recristalización partiendo de una solución acuosa. Es evidente, pues, que hay agua libre en la microestructura del cemento, ya que es necesaria para proporcionar el medio en el que la fase estable pueda nuclearizarse y crecer. Esta conclusión sugiere por qué una alta proporción de agua a cemento, en la mezcla original, puede facilitar la pérdida de solidez en un cemento de alta alúmina.

¿Pueden inhibirse las transformaciones? ¿No podría lograrse que el producto estable se produjese en primer lugar? Es difícil ver cómo podría evitarse el cambio, salvo por la práctica aceptada de hacer la mezcla relativamente seca, de tal manera que la cantidad de agua disponible para promover esta transformación no deseable sea mínima. Tal vez si la hidratación inicial se llevase a cabo a una temperatura suficientemente alta, el hidrato estable podría desarrollarse más rápidamente. Las temperaturas necesarias son probablemente impracticables por ser demasiado altas. Pero, en cualquier caso, es improbable que una matriz de cemento compuesta de poliedros facetados desarrolle la red interpenetrada que parece ser un requisito para conseguir una alta solidez.

En miembros estructurales, tales como una viga de hormigón pretensado prefatigado, hay otros problemas prácticos. Aunque se acepta que la mejor aproximación consiste en utilizar la mínima cantidad de agua durante la mezcla del cemento, ¿cúal puede ser la uniformidad de una mezcla como ésta en una sección de 25 a 30 centímetros de ancho y de 5 a 10 metros de longitud? Es instructivo considerar un ejemplo.

Se cortó una sección transversal de dos centímetros de espesor de una viga de cemento de alta alúmina de 6 metros. La viga tenía una inserción de madera en la parte posterior, quizá para que se pudiesen clavar objetos sobre ella. La sección transversal se dividió luego en piezas más pequeñas y se determinó la conversión relativa (de componente metaestable a estable) para cada pieza, mediante un análisis termoquímico.

Tal como podía esperarse de una estructura tan poco homogénea como ésta, la distribución era variable, con algunas regiones de conversión baja cerca de las superficies y alrededor de la madera insertada, zona en que el hormigón se había secado localmente por la absorción de agua. Es inevitable suponer que las variaciones en la otra dirección, a lo largo de la longitud de la viga, mostrarían una amplitud mayor. Esto permite apreciar los problemas prácticos que aparecen al querer asegurar homogeneidad en un material como el que tratamos, dentro de unos límites precisos y para grandes dimensiones. Estas variaciones de los valores de conversión han de reflejarse también en variaciones de la dureza mecánica del hormigón de uno a otro punto. La dureza de una viga determinada dependerá de su parte más débil.

Uno de los rasgos característicos de los cementos (y de los hormigones en general) es que pueden desarrollar una gran fuerza compresiva durante la hidratación, mientras que su resistencia a la tracción permanece muy baja. Por esta razón suele ser necesario reforzar las vigas de hormigón o someterlas a un prefatigado o postfatigado de compresión por medio de la inclusión de hilos o varas de acero estirado. En general, el diseño de ingeniería de una construcción tiene que ser tal que el hormigón no se vea sometido a una fatiga de tensión indebida. Se puede entender fácilmente que cualquier mejora que se consiguiese en las propiedades de tensión de un cemento o de un hormigón tendría una importancia de alcance universal en la práctica constructiva.

Si los procesos de fraguado y de endurecimiento dependen de la interpenetración y entrelazado de las fibrillas gelatinosas o cristales facetados, como sugieren los recientes hallazgos que hemos descrito, es razonable preguntarse si no podría mejorarse de alguna manera la resistencia a la tracción poniendo mayor énfasis en el desarrollo de las características del entrelazado, tal vez provocando un entretejido más compacto entre los productos de crecimiento. La modificación de la microestructura de los metales ha tenido una importancia fundamental en la tecnología metalúrgica y también se aplica ya en la solidificación de muchas aleaciones, pero no con el cemento.

Respecto de los hidratos cristalinos bien desarrollados del cemento de alta alúmina, las esperanzas de modificaciones útiles son inciertas. Con las fibras irregulares del cemento portland, sin embargo, podrían esperarse algunas mejoras si, por ejemplo, las fibras pudieran hacerse menos regulares y más intrincadamente entretejidas. Es instructivo considerar las largas, delgadas y densamente entrelazadas fibrillas que crecen de los cristales del sulfato ferroso. Es evidente que sería valioso averiguar si no es posible lograr algo semejante con el cemento portland, y, de ser así, cómo podría controlarse esta curiosa forma de crecimiento.

Aniones de metales alcalinos

Los metales alcalinos, tales como el sodio, resultan familiares en forma de cationes: iones positivos que se forman cuando los átomos pierden un electrón. Se ha encontrado, recientemente, que pueden ganar un electrón para formar aniones muy reactivos

James L. Dye

Los metales alcalinos forman un grupo de elementos cuya propiedad más notable y familiar es su facilidad para desprenderse de un electrón. El metal alcalino sodio, por ejemplo, da un electrón al cloro para formar el cloruro sódico. El sodio metálico está tan predispuesto a desprenderse de un electrón, que llega incluso a partir las moléculas de agua, desplazando un átomo de hidrógeno y formando hidróxido sódico: la reacción puede resultar muy violenta. Debido a que el sodio es tan reactivo en forma metálica, no se le encuentra así en la naturaleza, sino solamente como ion positivo, o catión, que se representa por Na^+. El catión se forma en cuanto el átomo cede uno de sus electrones a algún átomo de otra especie química.

Estas propiedades de los metales alcalinos han sido conocidas desde los inicios de la ciencia química. Por ello ha sido tanto más sorprendente el descubrimiento de que los metales alcalinos también pueden aceptar un electrón, actuando de manera exactamente opuesta a la usual. La adición de un electrón a un átomo neutro de sodio, por ejemplo, produce el ion cargado negativamente, o anión, Na^-. Se sabía, desde hace algún tiempo, que este anión era estable en el sodio gaseoso; recientemente, se ha descubierto que también se encuentra en disoluciones e incluso como componente de una sal cristalina. También parece que pueden prepararse los aniones de todos los demás metales alcalinos; se han observado ya el del potasio (K^-), el del rubidio (Rb^-) y el del cesio (Cs^-). Quizá llegue a resultar posible el preparar sales con el anión más simple imaginable: el electrón aislado.

Curiosamente, la clave para preparar los aniones de los metales alcalinos reside en conseguir atrapar los cationes alcalinos dentro de una molécula orgánica con una estructura tipo jaula. Normalmente, cualquier ion metálico negativo presente en la solución reaccionaría rápidamente con los iones positivos para dar átomos neutros del metal. Cuando los cationes se encuentran secuestrados dentro de moléculas orgánicas tipo jaula, el complejo resultante es tan estable que impide tal reacción. Los iones negativos y los cristales que les contienen son, sin embargo, altamente reactivos. No pueden ser expuestos al aire o a la humedad, y sólo resultan estables durante largos períodos de tiempo si se guardan a baja temperatura.

El grupo de los metales alcalinos comprende seis elementos: litio, sodio, potasio, rubidio, cesio y francio. (El último es una especie radiactiva.) En la disposición habitual de la tabla periódica de los elementos, los metales alcalinos aparecen en la primera columna, debajo del hidrógeno. Esta clasificación refleja las similitudes en las propiedades químicas de todos los metales alcalinos y en las estructuras electrónicas de sus átomos: cada uno tiene un electrón solitario en la capa de electrones más externa, la capa de valencia.

En cualquier átomo neutro debe haber los electrones necesarios para compensar exactamente la carga de los protones del núcleo. Los electrones tienden, por otra parte, a disponerse en la configuración que presente la mínima energía, pero esto pueden hacerlo solamente dentro de los límites permitidos por las reglas de la mecánica cuántica, que confinan los electrones dentro de los "orbitales". En física clásica, un orbital podría considerarse como la trayectoria de un electrón. La interpretación más precisa de la mecánica cuántica consiste en considerar que el orbital define la probabilidad de encontrar el electrón en una posición dada. De manera equivalente, se puede considerar el orbital como la representación de la densidad de una carga electrónica distribuida en el espacio.

La secuencia en que se llenan los orbitales disponibles viene determinada por una regla que prohíbe que dos electrones cualesquiera de un átomo ocupen exactamente el mismo estado. El primer orbital que se llena se denomina orbital $1s$; los electrones en el orbital $1s$ están distribuidos en un pequeño volumen esférico centrado en el núcleo. Esta distribución es completamente simétrica y no define ninguna dirección preferencial en el espacio. El orbital $1s$ sólo puede estar ocupado por un electrón o por dos electrones; en este último caso, los dos electrones tienen valores opuestos de momento angular de espín. (Según la mecánica cuántica, el electrón debe girar perpetuamente alrededor de su eje, y solamente existen dos orientaciones posibles, en el espacio, para este giro.)

El segundo orbital que se llena se denomina $2s$. Tiene la misma forma que el primero, pero, dado que los dos electrones que pueden ocuparlo tienen una energía superior, se hallan dispersos en una esfera de mayor diámetro.

Llegados a este punto de la construcción del átomo, es cuando se encuentran los primeros orbitales no esféricos. Se llaman orbitales tipo p, y son posibles tres de ellos; los tres reciben la denominación $2p$. Cada uno de estos orbitales $2p$ está constituido por dos lóbulos, uno a cada lado del núcleo, y están distribuidos en el espacio de tal manera que cada orbital es perpendicular a los otros dos. Los seis electrones que llenan los orbitales $2p$ (a pares, con espines opuestos) tienen un momento angular orbital unidad además de su momento angular de espín. En otras palabras, además de girar alrededor de su eje, tienen un movimiento de rotación alrededor del núcleo.

Para llenar los orbitales $1s$, $2s$ y $2p$ ha-

cen falta, pues, 10 electrones. Si se añaden más electrones al átomo, ocuparán primero el orbital 3*s* y después los tres orbitales 3*p* posibles. Estos orbitales son de forma similar a los correspondientes 2*s* y 2*p*, pero tienen una extensión mayor. Los átomos más pesados tienen aún otros tipos de orbitales (denominados *d, f* y *g*) con geometrías más complicadas y con valores más elevados del momento angular orbital.

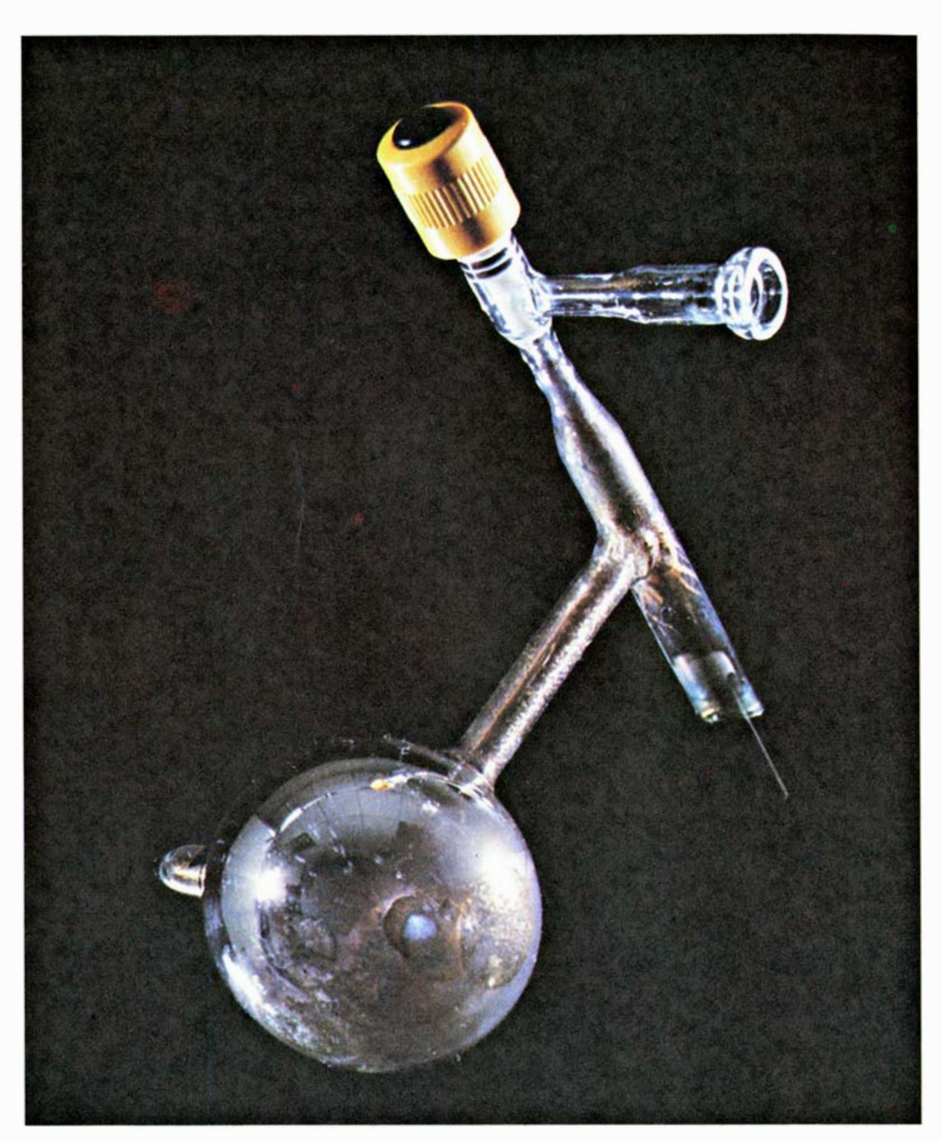
Resulta de una importancia fundamental en química el hecho de que un átomo que tenga todos sus orbitales *s* y *p* completos, posee una estabilidad excepcional. Estos átomos son los que constituyen la serie de los elementos denominados gases nobles. El primero de ellos es el helio, que posee solamente dos electrones, que llenan el orbital 1*s*. El neón, con 10 electrones, tiene los orbitales 1*s*, 2*s* y 2*p* completos; el argón tiene todos los orbitales completos hasta los 3*s* y 3*p*. La extracción o la adición de un electrón a los átomos de estos elementos resulta muy difícil.

Los metales alcalinos pueden considerarse como gases nobles con un electrón adicional (lo que requiere también, como es lógico, la adición de un protón en el núcleo). El litio tiene el orbital 1*s* lleno del átomo de helio más un electrón adicional que llena la mitad del orbital 2*s*. El sodio tiene los orbitales 1*s*, 2*s* y 2*p* llenos, como el neón, y además tiene un electrón en el orbital 3*s*. Los restantes metales alcalinos son de estructura similar; en todos ellos, el electrón más exterior se encuentra solitario en un orbital tipo *s*.

La fuerte tendencia de los metales alcalinos a perder un electrón y formar un catión puede comprenderse fácilmente dentro de este esquema. Con la pérdida del electrón de valencia, cada uno de los metales alcalinos adquiere la estructura extraordinariamente estable del gas noble correspondiente. A pesar de que el ion resultante tiene una carga positiva no compensada, en presencia de un aceptor de electrones resulta más estable que el propio átomo neutro.

Al añadir un electrón a un metal alcalino, resulta una configuración que es mucho menos favorable que la configuración de gas noble del catión. Sin embargo, el electrón extra permite llenar el orbital tipo *s*, y un orbital lleno es algo más estable que otro semilleno. Bajo ciertas circunstancias cuidadosamente preparadas, este pequeño aumento en estabilidad es suficiente para favorecer la existencia de aniones de los metales alcalinos.

La química de las especies atómicas



LOS ANIONES DE SODIO son un componente esencial de un líquido azulado y de una película de color de bronce formados en el interior de un recipiente de vidrio en el que se ha hecho el vacío. Para prepararlos, se depositó primero una capa de sodio metálico en la superficie interior del globo situado a la izquierda. Parte de este sodio se disolvió luego en etilamina. Los iones positivos de sodio (cationes) quedan atrapados en la disolución por una molécula orgánica, llamada un "cryptand", que tiene una estructura parecida a una jaula. Los iones negativos (aniones) están dispersos en el solvente y dan a la disolución el fuerte color azul que puede verse al fondo, a la derecha. La película, que precipita sobre las paredes del recipiente cuando se evapora parte del solvente, está formada por aniones y cationes atrapados. El objeto de plástico *(arriba)* es una válvula de vacío.



CRISTALES DE UNA SAL, compuestos de aniones de sodio y de un complejo de cationes de sodio y de moléculas "cryptand". Como cualquier otra sal, ésta se forma mediante la donación de un electrón por parte de una especie química a otra especie, creando un par de iones. Lo que es nuevo en este caso es que el donante y el receptor de los electrones son formas del mismo elemento. Los cristales precipitaron de una solución al enfriarla. Su brillo metálico es engañoso. La sal no es un metal, sino un semiconductor. Los cationes de sodio son muy reactivos. La sal no puede exponerse al aire; incluso en el vacío se descompone lentamente a menos que se almacene a baja temperatura.

ionizadas se estudia mejor cuando se encuentran en solución. Para poder estudiar así los metales alcalinos, precisamos una sustancia capaz de disolverlos. La serie de investigaciones que acaba de culminar con la observación de aniones de metales alcalinos empezó en Alemania en el año 1864, con el descubrimiento hecho por W. Weil de que los metales alcalinos se disuelven en amoníaco líquido anhidro (NH_3). Se obtienen así disoluciones con una coloración azul oscuro.

El proceso por el cual se disuelve una sustancia iónica consiste en una interacción eléctrica íntima entre las moléculas del disolvente y los iones del sólido. El disolvente debe, en general, ser un líquido polar, es decir, debe estar constituido por moléculas que, aun siendo eléctricamente neutras, posean regiones eléctricamente cargadas. El amoníaco es un ejemplo: en el estado líquido, la molécula de amoníaco tiene una estructura tal que los tres átomos de hidrógeno están agrupados a un lado del átomo de nitrógeno. El nitrógeno tiende a adquirir una carga eléctrica negativa, compensada por una carga positiva que aparece sobre los hidrógenos, y como resultado la molécula es un dipolo eléctrico: tiene un extremo negativo y otro positivo.

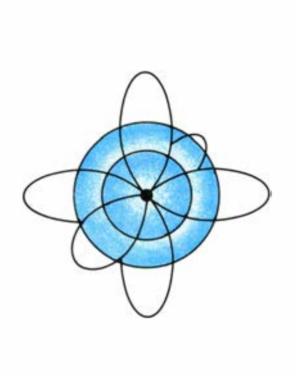
Cuando se disuelve una sal, como el cloruro sódico, se descompone en io-

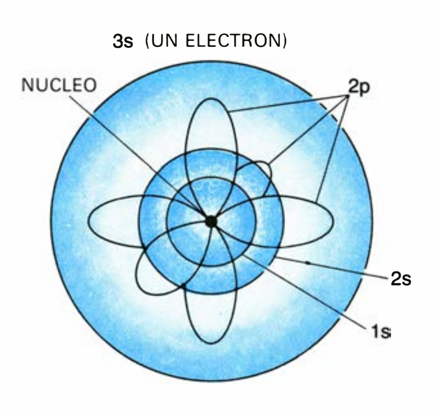
GASES NOBLES

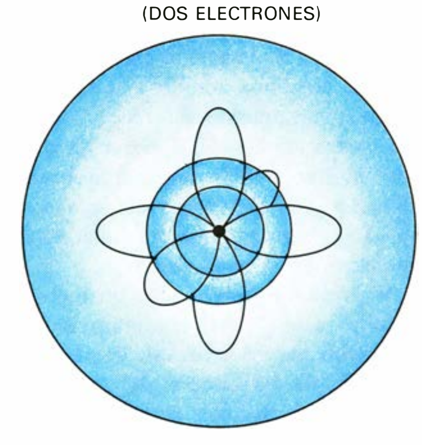
H HIDROGENO																	He HELIO
Li LITIO	Be											B	C	N	O	F	Ne NEON
Na SODIO	Mg											Al	Si	P	S	Cl	Ar ARGON
K POTASIO	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr CRIPTON
Rb RUBIDIO	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe XENON
Cs CESIO	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn RADON
Fr FRANCIO	Ra	Ac															

GASES NOBLES

LOS METALES ALCALINOS son los elementos situados debajo del hidrógeno en la primera columna de la tabla periódica. Se agrupan así debido a que, al tener estructuras electrónicas similares, poseen propiedades químicas parecidas. En cuanto a su configuración electrónica, los átomos de los metales alcalinos pueden considerarse como atomos de un gas noble con un electrón adicional. Los gases nobles (situados en la columna más a la derecha) son los elementos más estables y por ello los metales alcalinos ceden fácilmente un electrón para convertirse en un catión con la estructura electrónica de gas noble. Los átomos neutros de metal alcalino se desprenden con tanta facilidad del electrón sobrante, que incluso a veces llegan a cederlo a otro de los átomos alcalinos; a consecuencia de lo cual, podrá convertir este último en un anión alcalino.







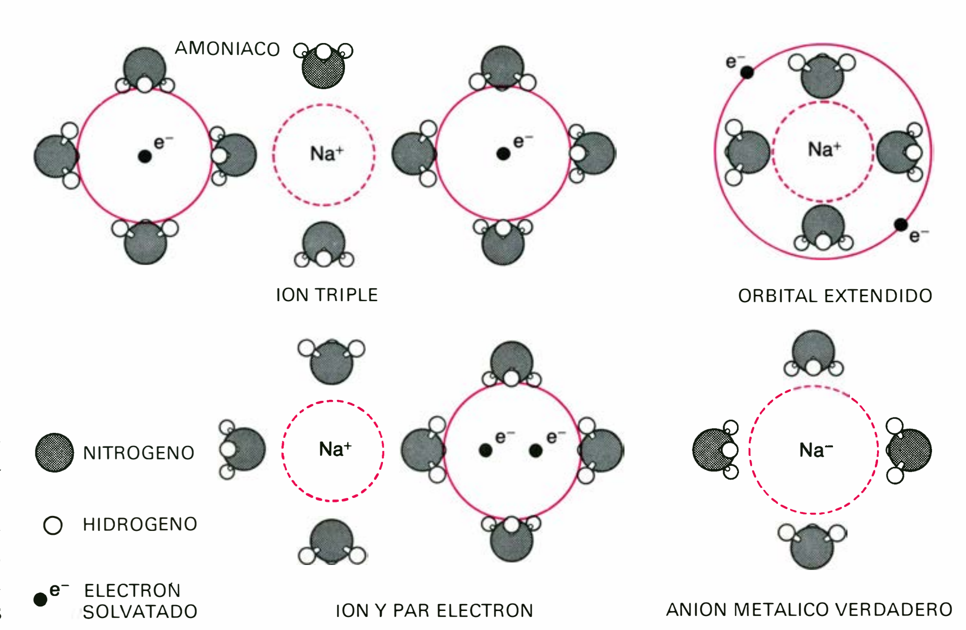
LA GANANCIA O LA PERDIDA DE UN ELECTRON altera profundamente las propiedades químicas de los metales alcalinos. Un átomo de sodio neutro *(centro)* tiene dos electrones en cada uno de los cinco orbitales interiores y un solo electrón en el más externo, en el llamado orbital de valencia. Los orbitales definen la distribución espacial de la carga electrónica, y pueden contener, cada uno, dos electrones como máximo. El orbital de energía más baja, denominado 1*s*, tiene forma esférica; el siguiente llamado 2*s* es también esférico, pero mayor. Hay, además, tres orbitales 2*p*, cada uno con dos lóbulos. Por último, está el orbital 3*s*, el más exterior, de forma esférica. En un catión de sodio *(izquierda)*, el electrón 3*s* se ha desprendido y el ion tiene la configuración electrónica del neón, con los orbitales 1*s*, 2*s* y 2*p* completos. El anión sodio posee un electrón suplementario *(derecha)*, que ha sido capturado, de manera que el orbital 3*s* se encuentra ya completo, esto es, con dos electrones.

nes positivos y negativos; en este caso Na^+ y Cl^-. Lo que mantiene estos iones disueltos es una interacción con las moléculas polares del disolvente. Por ejemplo, la carga positiva de uno de los iones Na^+ atrae hacia sí las regiones más negativas de las moléculas de disolvente, y queda rápidamente rodeada de una capa de moléculas orientadas. Los iones negativos atraen un revestimiento análogo de moléculas de disolvente, pero orientadas por supuesto en sentido inverso. Estas capas de moléculas de disolvente orientadas estabilizan la disolución e impiden a los iones agregarse para constituir una sal.

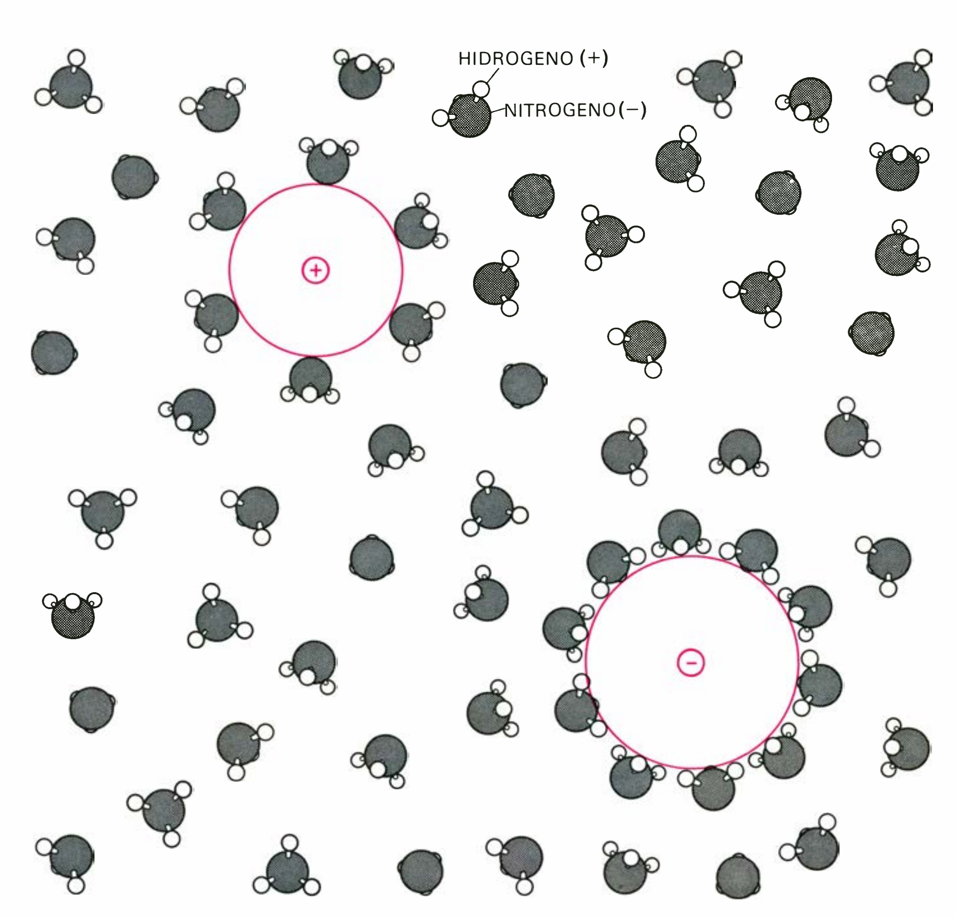
La mayoría de los metales se disuelven mucho más difícilmente que las sales, por la sencilla razón de que no es fácil formar aniones a partir de ellos. Los metales alcalinos se disuelven en el amoníaco debido a su extraordinaria disposición para perder un electrón. La disolución resultante contiene cationes metálicos, tales como el Na^+, indistinguibles de los que se forman en una disolución salina. En cambio, los iones negativos son muy especiales: son simplemente electrones que han sido liberados en la disolución. Estos electrones se llaman electrones solvatados. Se comportan como cualquier especie cargada negativamente: atraen y orientan las moléculas polares de disolvente, pero con una movilidad considerablemente mayor, debido a que son miles de veces más ligeros que cualquier otro ion.

¿Cómo podemos saber qué especies químicas están presentes en una disolución? ¿Cómo se puede demostrar, por ejemplo, que las disoluciones de metales alcalinos en amoníaco contienen electrones disueltos y no algún otro ion negativo, quizá verdaderos aniones del metal alcalino? Existen numerosas técnicas aplicables a estos problemas, pero vamos a considerar sobre todo dos de ellas. Estas técnicas caracterizan los iones presentes en disoluciones por sus propiedades magnéticas y por las longitudes de onda de la radiación que absorben, respectivamente.

El magnetismo de los electrones se debe a su momento angular. Por ejemplo, el giro de un electrón individual da lugar a un pequeño campo magnético, de forma muy parecida a como lo hace una corriente de electrones que circule por una espira. Sólo en algunos átomos pueden observarse estos campos. Recordemos que en los orbitales atómicos completos los electrones están apareados dos a dos, con espines opuestos: de ello resulta que los campos magnéticos que crean tienen también direcciones opuestas y se cancelan exactamente. En un átomo que



LOS IONES EN DISOLUCION no pueden agregarse para formar un sólido debido a su interacción eléctrica con las moléculas del disolvente. Para los metales alcalinos (y para otras muchas sustancias), sólo los líquidos con moléculas polares son buenos disolventes. Una molécula polar es aquella que posee regiones cargadas eléctricamente. Un ejemplo es el amoníaco: el átomo de nitrógeno en el amoníaco líquido tiende a adquirir una carga negativa, compensada por una carga positiva en los átomos de hidrógeno. Un ion sumergido en este líquido atrae hacia sí una capa de moléculas de disolvente orientadas. El amoníaco puede llegar a disolver todos los metales alcalinos.



LOS IONES NEGATIVOS que pueden darse en las disoluciones de metales alcalinos comprenden varias estructuras con carga negativa, además de los aniones del metal alcalino. Algunas de estas estructuras se han representado esquemáticamente para el caso del sodio disuelto en amoníaco. Un catión sodio (Na^+) puede atraer dos electrones rodeados de moléculas de disolvente, formando una estructura llamada ion triple. Otra estructura podría estar formada por la asociación de un catión con un par de electrones incluidos en una misma "jaula" de moléculas de disolvente. También podrían asociarse dos electrones a un catión en un "orbital extendido", que incluye varias moléculas de disolvente además del catión. Todas estas configuraciones se distinguen por el hecho de que las moléculas de disolvente se interponen entre el catión sodio y los dos electrones externos; en cambio, en un anión de metal alcalino, estos dos electrones en cuestión ocupan un orbital $3s$ normal.

no tenga orbitales incompletos, todos los campos se cancelan y no originan un magnetismo neto. Una sustancia constituida por átomos de este tipo se llama diamagnética: no tiene magnetismo propio y, cuando está sometida a un campo externo, se opone débilmente a él, repeliendo parcialmente las líneas de fuerza.

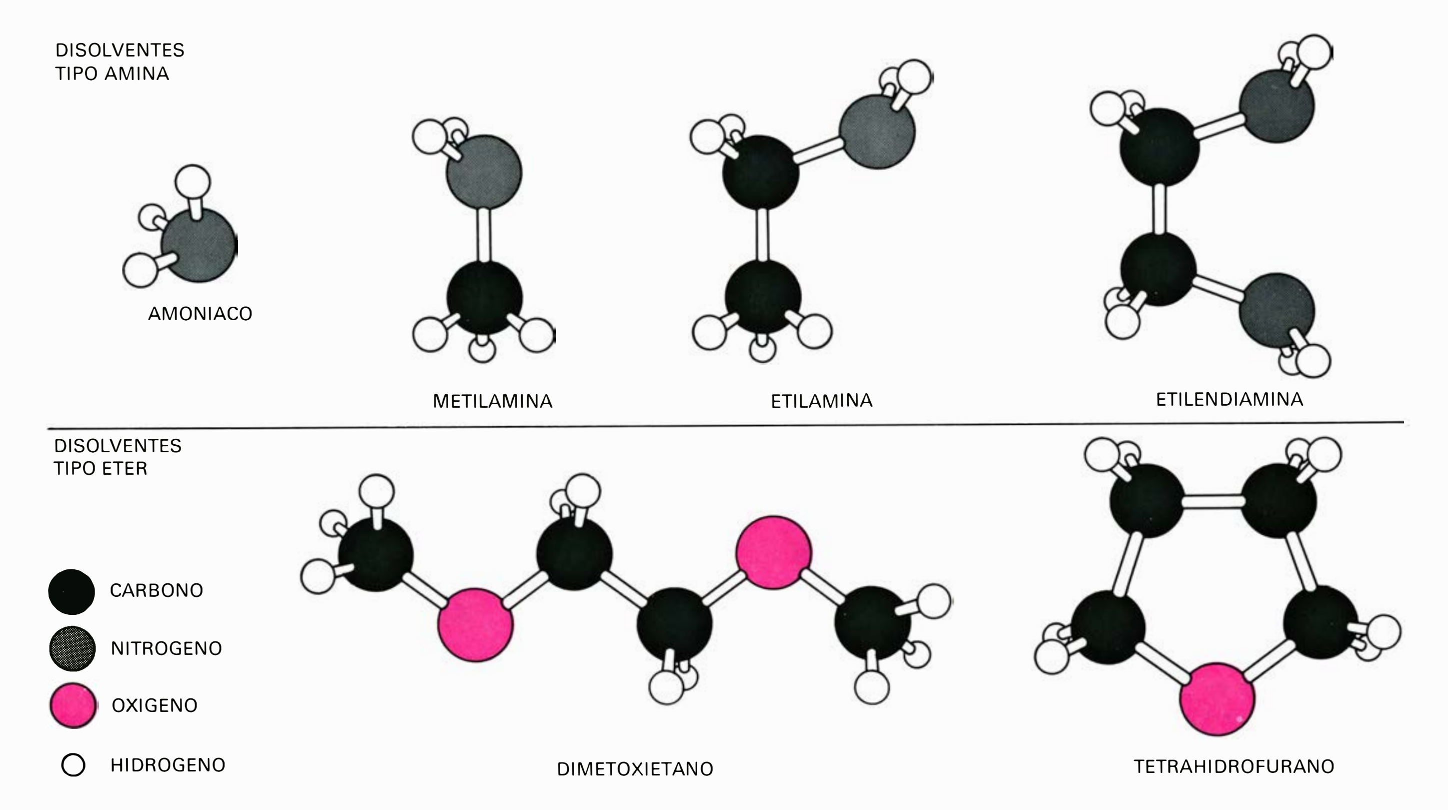
Los átomos con orbitales incompletos poseen espines electrónicos desapareados y pueden producir un campo magnético neto. Si este campo es muy fuerte, como ocurre en el hierro y en otros metales directamente relacionados con él, el material se llama ferromagnético. En un material ferromagnético los espines de los electrones desapareados se alinean espontáneamente, por lo menos en dominios microscópicos, y el material presenta un magnetismo "permanente". Sin embargo, en la mayoría de las sustancias magnéticas, y entre ellas los metales alcalinos, el campo asociado a cada átomo es mucho menor, demasiado pequeño para alinear paralelamente los espines electrónicos. No obstante, estos campos pueden hacerse paralelos, y la sustancia magnetizarse débilmente, mediante un campo externo intenso. Las sustancias que poseen estas propiedades se llaman paramagnéticas.

En una disolución amoníaco-metal alcalino, los electrones solvatados deberían comportarse como los electrones desapareados en un átomo y, por ello, hacer que la solución fuese paramagnética. Por otra parte, la intensidad del paramagnetismo debería ser proporcional a la concentración de electrones solvatados. Se ha demostrado que estas disoluciones son realmente paramagnéticas, y es indiscutible que contienen electrones *solvatados.* No obstante la intensidad del campo inducido no es proporcional a la concentración, y de ello se deduce que también están presentes otras especies iónicas. La naturaleza de las desviaciones de la proporcionalidad indica que algunos electrones forman parejas constituyendo así especies diamagnéticas.

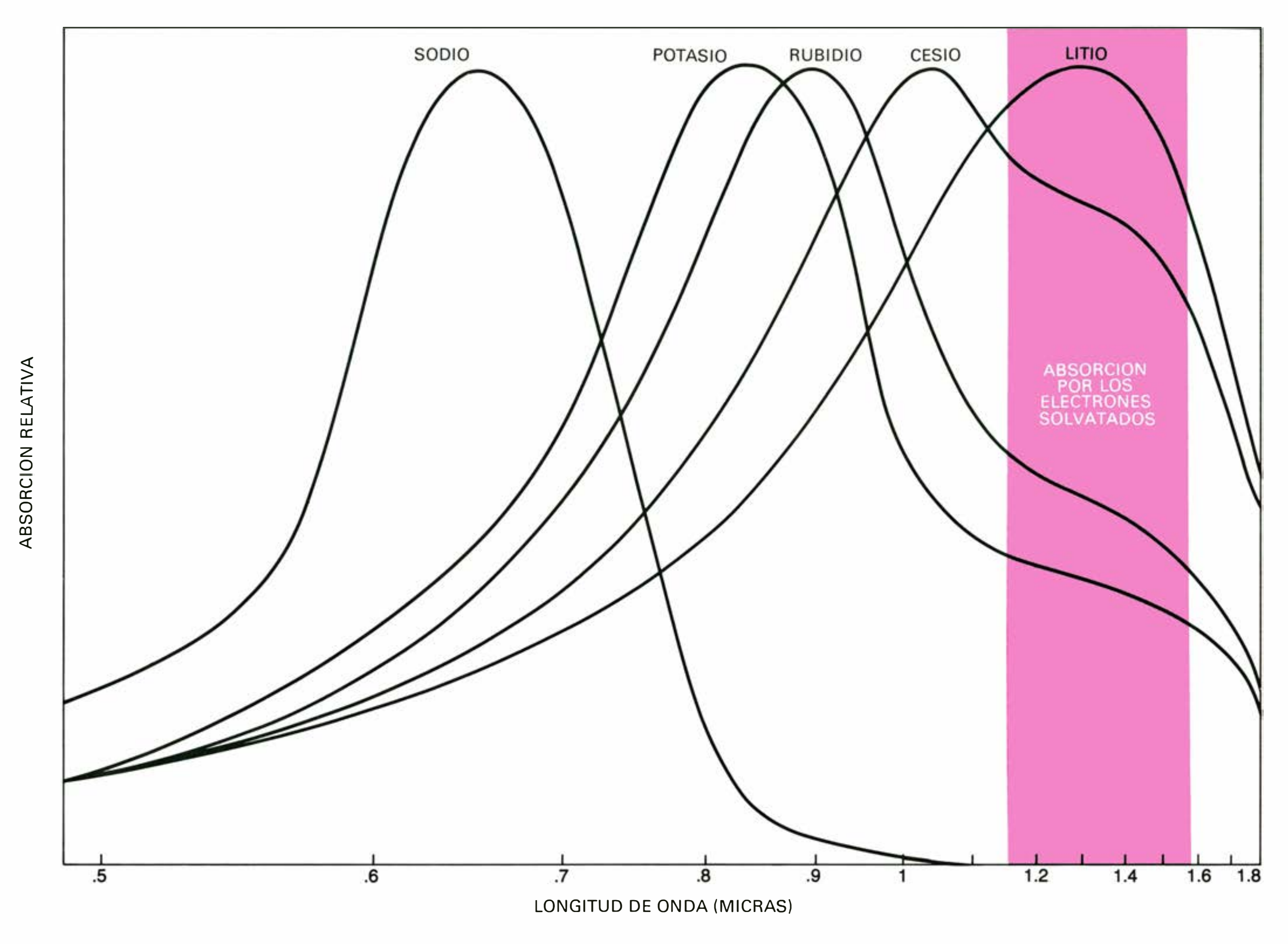
La naturaleza de los iones diamagnéticos en las disoluciones amoníaco-metal ha sido un tema de discusión durante décadas, y todavía no está resuelto. Un candidato evidente es el anión del metal alcalino, ya que un metal alcalino que capta un electrón suplementario tiene todos sus orbitales llenos, y debe ser diamagnético. La primera hipótesis de este tipo fue formulada en 1953 por el investigador alemán Werner Bingel, aunque la estructura en la que él pensaba no era un verdadero anión de metal. El imaginaba más bien que un catión metálico ordinario podría atraer hacia sí un par de electrones ligados entre sí. La posibilidad de que las disoluciones amoníaco-metal pudieran contener verdaderos aniones del metal alcalino fue discutida por primera vez en 1965 por Thomas R. Tuttle, Jr., Sidney Golden y Charles Guttman, de la Universidad de Brandeis.

Está claro que pueden existir aniones metálicos en las disoluciones amoniacales, pero hay que decir que no existe una prueba explícita de su presencia. Es más, el espectro de absorción de estas disoluciones aporta argumentos, aunque algo indirectos, para afirmar que el anión no puede ser la única especie diamagnética.

El color azul intenso de las soluciones amoníaco-metal es el resultado de una fuerte absorción de la luz en el extremo rojo del espectro visible y en la región del infrarrojo. Esta absorción de las longitudes de onda largas es común en los materiales que contienen electrones que no están localizados sobre átomos particulares: los metales son un buen ejemplo de ello. En las soluciones amoniacales, la luz y la radiación infrarroja se absorben tanto por los electrones disueltos como por las especies diamagnéticas. Es significativo que el espectro de absorción no se vea afectado por la elección del metal alcalino de la disolución. Si las especies diamagnéticas fueran los aniones de metal alcalino, el espectro sería algo distinto para cada uno de ellos: es evidente que los electrones interaccionan fuertemente sólo con otros electrones y no con los iones metálicos.



LOS DISOLVENTES apropiados para los metales alcalinos constan de dos clases principales de compuestos: el amoníaco y sus derivados (las aminas), que contienen nitrógeno, y los éteres, que contienen oxígeno. Las moléculas de estos líquidos son suficientemente polares para disolver los metales; además no reaccionan químicamente. Exceptuando el amoníaco, con estos disolventes sólo se obtienen concentraciones moderadas.



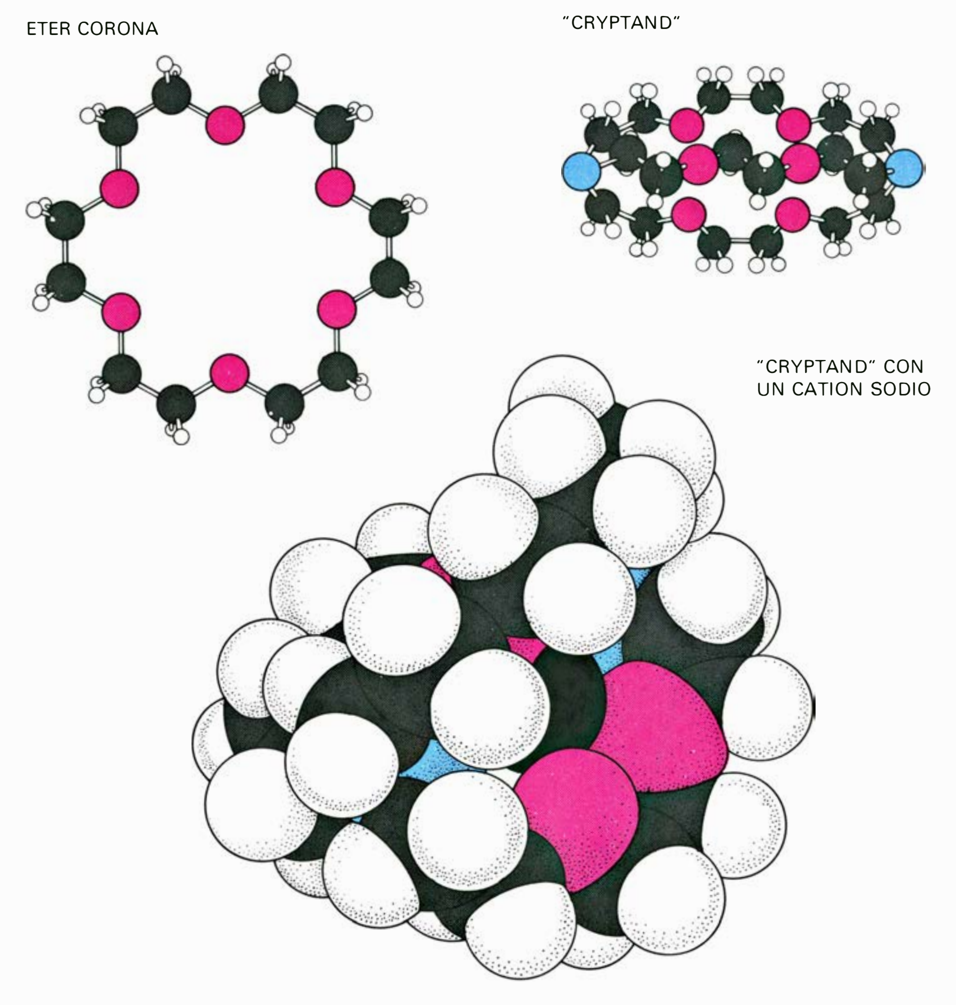
ABSORCION DE LA LUZ por disoluciones de metales alcalinos en aminas y en éteres, que sugiere la presencia, en la disolución, de dos especies distintas responsables de la absorción. A longitudes de onda de unas 1,4 micras, aproximadamente, en la región infrarroja del espectro, se observa un pico de absorción que ha sido atribuido a los electrones solvatados. La longitud de onda correspondiente a este pico no depende del metal presente en la disolución. A longitudes de onda más cortas, aparece un segundo pico que se atribuye a la absorción por los aniones de metal alcalino; es significativo que este pico aparezca a longitudes de onda distintas para cada metal. En las disoluciones de potasio, rubidio y cesio se observan ambos picos. En cambio, en las disoluciones de litio parece ser que no están presentes los aniones del metal, ya que en el espectro sólo se observa el pico asociado a los electrones solvatados. Por otra parte, el sodio parece formar aniones tan rápidamente que no aparecen electrones solvatados. Los espectros representados en esta figura se han obtenido con disoluciones de metales alcalinos en etilendiamina.

Es posible que no existan aniones metálicos en las soluciones amoniacales, debido a que el amoníaco es un disolvente demasiado bueno. El catión puede estar tan fuertemente estabilizado por su interacción con las moléculas polares del amoníaco, que no pueda coexistir con ninguna otra forma del metal. Cualquier anión que pudiera formarse tendería a romperse en dos electrones solvatados y un catión.

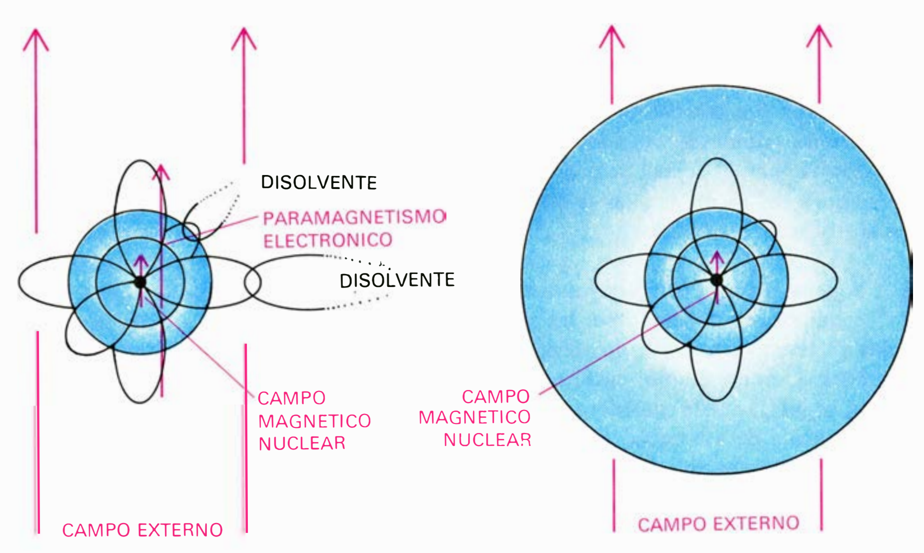
La elección de posibles disolventes para los metales alcalinos es limitada. La mayoría de los líquidos no son suficientemente polares como para disolver los metales: los hidrocarburos simples, tales como las fracciones del petróleo o el benceno, por ejemplo, son inertes frente a los metales alcalinos. Por otra parte, muchos de los disolventes polares reaccionan con estos metales formando compuestos con ellos. Ya se ha mencionado antes la reacción del metal sodio con el agua (un disolvente muy polar). Reacciones similares, incluso algunas de ellas violentas, dejan fuera de consideración a los alcoholes, las acetonas y los hidrocarburos clorados.

Un grupo de disolventes de los metales alcalinos está formado por las moléculas relacionadas con el amoníaco. La metilamina, por ejemplo, tiene por fórmula CH_3NH_2; puede considerarse como una molécula de amoníaco en la que un átomo de hidrógeno ha sido reemplazado por un grupo metilo ($-CH_3$). El litio se disuelve rápidamente en metilamina líquida, aunque el sodio, el potasio, el rubidio y el cesio son mucho menos solubles. Otro compuesto relacionado con el amoníaco es la etilendiamina, de estructura $NH_2–CH_2–CH_2–NH_2$. Podemos imaginar esta molécula como dos moléculas de metilamina unidas entre sí. En mi laboratorio de la Universidad del Estado de Michigan, mis colegas y yo hemos encontrado que la etilendiamina disuelve bien los metales alcalinos.

Además de las aminas simples, el otro grupo importante de disolventes de los metales alcalinos son los éteres, el tipo de moléculas orgánicas que incluyen la estructura $–CH_2–O–CH_2–$. El éter más familiar, empleado como anestésico, es el éter dietílico($CH_3–CH_2–O–CH_2–CH_3$), pero los metales alcalinos no se disuelven en este líquido. Los éteres más efectivos como disolventes son los que contienen más de un átomo de oxígeno (poliéteres), tales como el dimetoxietano ($CH_3–O–CH_2–CH_2–O–CH_3$). Otro disolvente es el tetrahidrofurano, un éter de estructura cíclica. Debemos señalar que ninguna de estas sustancias es un gran disolvente de los metales alcalinos. Con ellas no pueden prepararse soluciones concentradas, pero puede disolverse una cantidad de metal suficiente para obser-



LOS ETERES CICLICOS Y LOS "CRYPTAND" envuelven los cationes de los metales alcalinos mejorando su solubilidad y facilitando indirectamente la formación de aniones de metal alcalino. Las estructuras de estos dos tipos de moléculas se muestran esquemáticamente en la parte superior; los "cryptand" resultan más eficaces debido a que su cavidad tridimensional aísla mejor el ion frente al disolvente. Una representación tridimensional del "cryptand" muestra que el catión sodio encuentra un ajuste compacto en la cavidad de la molécula. En disoluciones y en sales, los "cryptand" estabilizan los cationes de los alcalinos, impidiendo su recombinación con los aniones.



LAS PROPIEDADES MAGNETICAS DE UN ATOMO dependen de su estado de ionización y del entorno químico. El campo magnético nuclear puede medirse (mediante la técnica llamada espectroscopía de resonancia magnética nuclear), y estas medidas pueden proporcionar una información indirecta del estado de ionización de los electrones atómicos. En los átomos o iones que tienen sus electrones externos en orbitales tipo *p (izquierda)*, las interacciones con las moléculas de disolvente tienden a intensificar el efecto producido por el campo magnético externo sobre el campo nuclear; a este fenómeno se le llama paramagnetismo. Si los electrones externos están en orbitales tipo *s (derecha)*, esta intensificación paramagnética no se presenta. Por espectroscopía se puede distinguir entre los aniones y los cationes contenidos en las disoluciones de metales alcalinos.

var las propiedades de las especies químicas formadas.

Las propiedades ópticas de estas disoluciones son significativamente distintas de las de disoluciones metal-amoníaco. La banda de absorción principal, en la región infrarroja, asociada a los electrones disueltos, puede observarse todavía, pero hay también otra banda de absorción bien definida a longitudes de onda más cortas. Lo más importante es que esta nueva banda de absorción tiene posiciones diferentes para cada uno de los metales alcalinos. En 1969, S. Matalon, Golden y Michael Ottolenghi sugirieron que esta nueva banda de absorción se debe a aniones del metal. Se ha comprobado ya que su hipótesis era correcta.

Es fácil demostrar que una solución contenga algunas especies con una carga eléctrica neta de - 1, pero un anión metálico genuino no sería la única estructura posible con esta propiedad. Por ejemplo, estas especies podrían ser un catión metálico unido débilmente a dos electrones solvatados independientes. Otra posibilidad sería que los dos electrones ocupasen un "orbital extendido" con el catión metálico en su centro, como en un ion verdadero, pero incluyendo también algunas moléculas de disolvente. O bien, un catión podría estar asociado a un par de electrones fuertemente ligados entre sí, pero sólo ligeramente ligados al ion metálico.

Todas estas estructuras pueden distinguirse inequívocamente del anión metálico verdadero. Se caracterizan por la presencia de un catión metálico sumergido en moléculas de disolvente, orientadas igual que las moléculas que rodean un ion positivo; los dos electrones están por alguna parte, fuera de una capa de moléculas de disolvente. En el anión metálico verdadero los dos electrones están fuertemente ligados al metal y el disolvente interacciona con el ion como una sola entidad, cargada negativamente.

Las medidas de efectos ópticos y magnéticos con disoluciones de metales alcalinos en éteres y aminas proporcionan datos que sugieren que los aniones metálicos verdaderos existen. En mi laboratorio, mis colegas y yo hemos mostrado que la banda de absorción óptica es debida a dos electrones equivalentes o intercambiables; este es el caso de los iones metálicos, y también de otros modelos. Con Leon M. Dorfman, de la Universidad del Estado de Ohio, hemos medido la velocidad de formación de un anión sodio, Na^-, a partir de electrones solvatados y cationes de sodio. Otros investigadores mostraron que las espe-

cies cargadas negativamente pueden romperse en un catión y dos electrones solvatados, que se recombinan de nuevo posteriormente. Hacia 1970 parecía verosímil que se habían descubierto los aniones metálicos, pero no podían excluirse otras interpretaciones. Uno de los impedimentos principales de un estudio más avanzado de estos iones era la baja solubilidad de los metales.

En 1970, Vincent A. Nicely, que trabajaba en mi laboratorio, sugirió que la solubilidad de los metales alcalinos podría mejorar con la ayuda de una nueva clase de compuestos, llamados éteres cíclicos, en corona. Estos éteres, desarrollados por Charles J. Pedersen de E. I. du Pont de Nemours and Company, habían probado ya que producían un gran incremento de la solubilidad de las sales alcalinas ordinarias. Como sugiere su nombre, tienen la forma de un anillo almenado. Los cationes encajan en la parte central del anillo, que tiene afinidad para las cargas positivas; el complejo puede disolverse rápidamente en un líquido polar, como una amina o un éter.

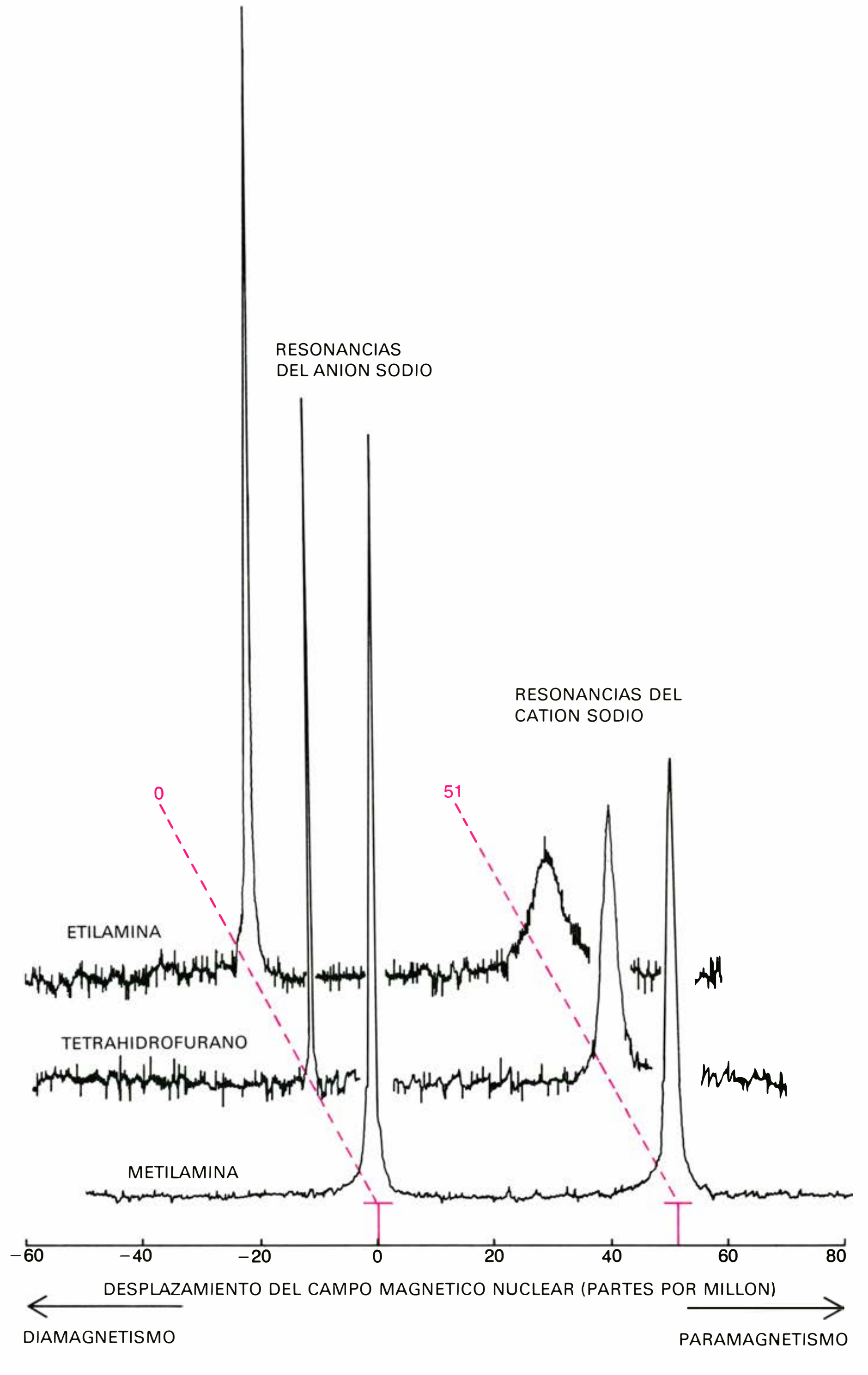
Químicamente, los éteres cíclicos no son demasiado distintos de los poliéteres, que habíamos empleado frecuentemente como disolventes, por lo que parecía razonable suponer que no reaccionarían fuertemente con los metales alcalinos. Esta suposición resultó ser correcta. Como esperábamos, las solubilidades aumentaron en varios órdenes de magnitud.

Otra clase de moléculas que forman complejos con los metales alcalinos fue estudiada posteriormente por Jean-Marie Lehn, de la Universidad Louis Pasteur de Strasbourg. Lehn llamó a estos agentes "cryptand", debido a que su estructura característica es en forma de una cripta tridimensional que incluye un catión, apantallándolo bien de las moléculas de disolvente del medio.

Todos los "cryptand" que nos interesan aquí tienen dos átomos de nitrógeno conectados por tres cadenas orgánicas más bien largas. Cada cadena es un éter con uno, dos o tres átomos de oxígeno. El simple hecho de designar el número de átomos de oxígeno en las cadenas sirve para identificar el "cryptand". El "cryptand" 1,1,1, con un oxígeno en cada cadena, tiene una pequeña cavidad que puede acomodar solamente un ion hidrógeno (un protón) o un catión litio Li^+. Los "cryptand" 2,2,1 y 2,2,2, con criptas más grandes, son adecuados para el ion sodio, Na^+.

Los "cryptand" pueden provocar un enorme aumento de la solubilidad. En etilamina pura, por ejemplo, el sodio es prácticamente insoluble. En presencia del "cryptand" 2,2,2, pueden disolverse varios gramos de sodio por litro de disolvente. La solubilidad se ha incrementado en un factor de 400.000 por lo menos.

La estabilidad del complejo metal-"cryptand" es la que hace posible estas solubilidades elevadas. Paradójicamente, la estabilidad del catión dentro de la cavidad de las moléculas de "cryptand" conduce también a una concentración elevada de aniones metálicos. De hecho, ajustando las cantidades de metal y de



ESPECTROS DE RESONANCIA MAGNETICA NUCLEAR, que proporcionan una prueba evidente de la existencia de verdaderos aniones metálicos en varias disoluciones de sodio. Las disoluciones concentradas necesarias para estas medidas se obtuvieron gracias a las moléculas de "cryptand". Como punto cero de la escala se elige el correspondiente al campo magnético no perturbado del núcleo de sodio en el catión gaseoso. Las desviaciones aparentes de este valor en las medidas del campo nuclear pueden ser debidas a campos electrónicos que contribuyen (paramagnetismo) o se oponen (diamagnetismo) al campo nuclear. Los picos relativamente anchos, con un desplazamiento paramagnético de 51 partes por millón, se asocian a los cationes confinados en el "cryptand". Los picos estrechos, no desplazados, son las resonancias producidas por los aniones sodio.

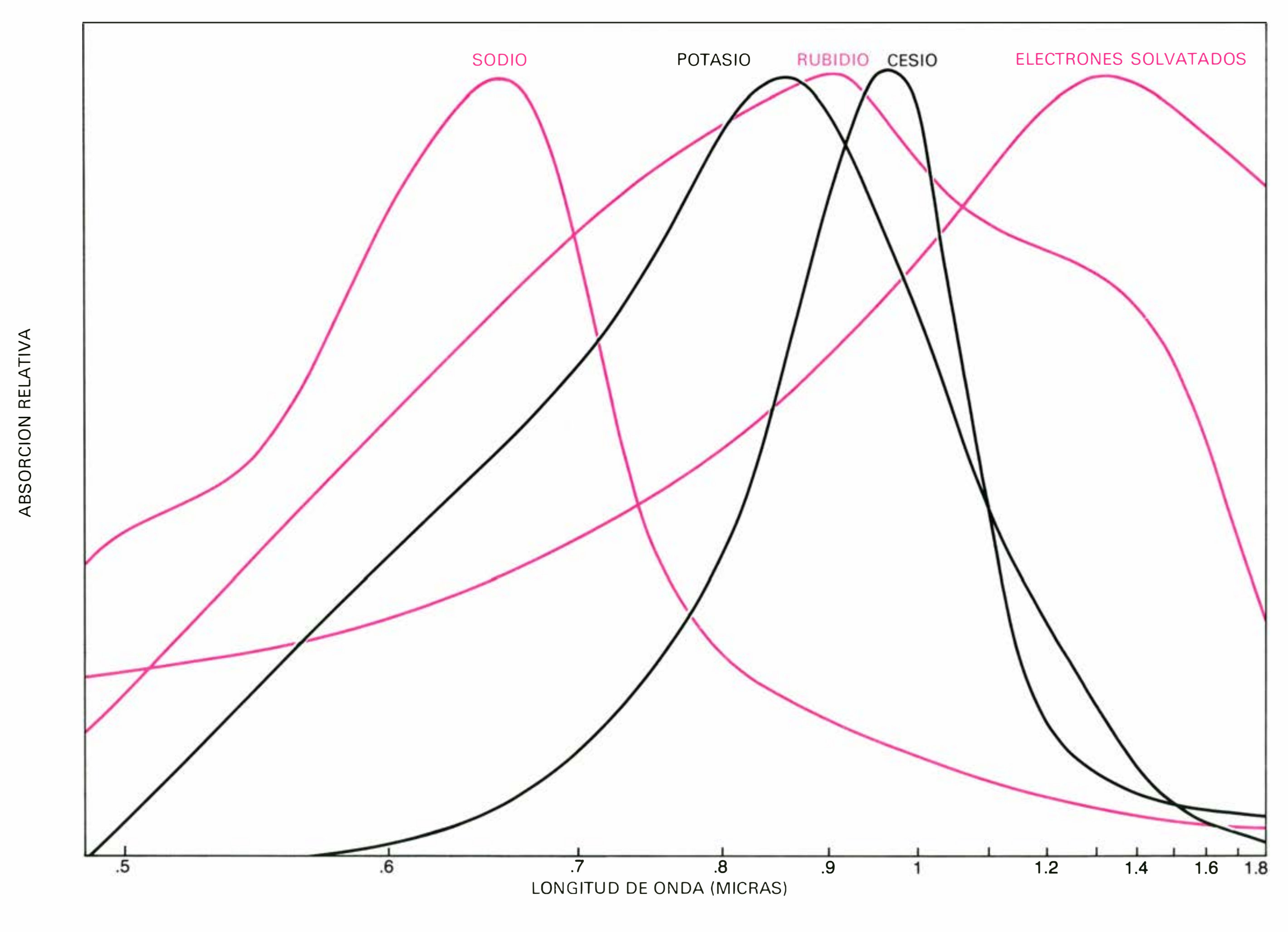
"cryptand", podemos controlar las especies iónicas presentes en la solución.

Cuando los átomos metálicos y las moléculas de "cryptand" están presentes en cantidades iguales, prácticamente todos los átomos metálicos se hallan en forma de cationes atrapados en las "jaulas" orgánicas; esta regla se cumple bien para el potasio, el rubidio y el cesio. La especie negativa predominante en estas soluciones es entonces el electrón "disuelto", ya que no hay átomos metálicos disponibles para formar aniones. Si el metal está presente en exceso, el catión atrapado constituye la especie positiva principal, pero ahora los átomos metálicos excedentes se combinan con los electrones para formar aniones del metal alcalino. Estas reglas generales son válidas para todos los metales alcalinos, exceptuando el litio y el sodio. (El francio no ha sido investigado todavía.) En aminas y en éteres, el litio parece ser reacio a formar el anión, sean cuales sean las condiciones. Por otra parte, el sodio forma un anión tan estable que, en sus soluciones, es raro encontrar electrones solvatados.

Con las disoluciones concentradas que pueden prepararse con los "éteres en corona" y con los "cryptand", es mucho más fácil estudiar las propiedades de los aniones de metales alcalinos. Por ejemplo, se puede utilizar la técnica conocida con el nombre de espectroscopía de resonancia magnética nuclear. Con esta técnica, podemos examinar aisladamente un elemento particular de la solución y deducir su estado químico.

Los núcleos atómicos pueden tener un momento angular de espín, de la misma manera que lo tienen los electrones, y estos núcleos pueden generar entonces campos magnéticos. Los núcleos de todos los metales alcalinos son magnéticos; el núcleo del sodio, por ejemplo, tiene cuatro estados de espín posibles (a diferencia de los dos estados posibles del electrón) y cuatro orientaciones posibles de su campo magnético respecto a un campo magnético externo. Este magnetismo nuclear no tiene ninguna influencia que se pueda detectar en las propiedades químicas del átomo, pero puede utilizarse convenientemente como sonda que permita analizar el campo magnético electrónico, el cual sí está afectado por el entorno químico del átomo.

En la práctica, el campo magnético nuclear se examina forzando al núcleo a cambiar su estado de espín. Puesto que cada estado de espín corresponde a una orientación distinta del campo magnético nuclear respecto a un campo externo, cada estado tiene una energía característica distinta. La magnitud de la diferencia de energías entre estados es proporcional a la intensidad del campo magnético externo que actúa sobre el núcleo. La energía necesaria para que el núcleo pase a un estado energético superior puede suministrarla un fotón, o cuanto de energía electromagnética; el núcleo puede caer de nuevo al estado de



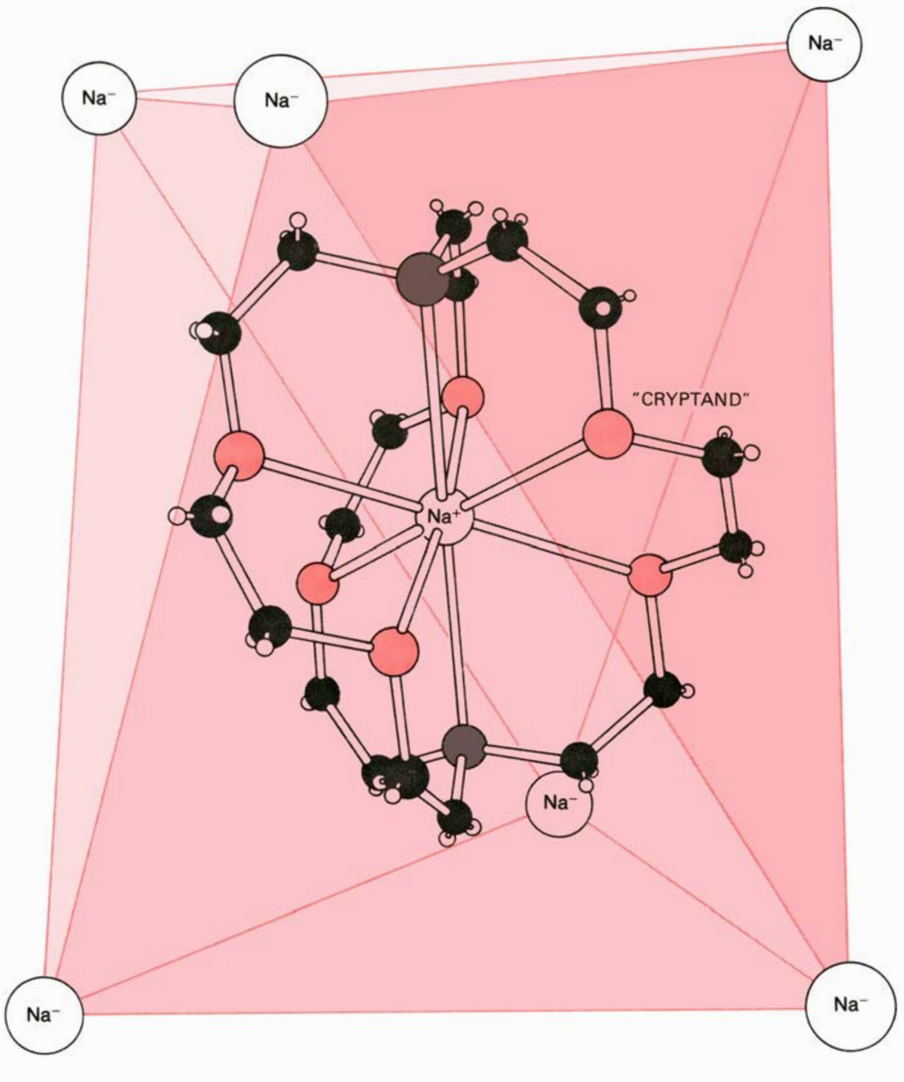
ESPECTROS DE ABSORCION obtenidos con películas sólidas; indican que tanto los aniones como los electrones solvatados no se alteran al evaporar el disolvente. Estas películas se preparan a partir de soluciones que contienen metal disuelto y "cryptand". El espectro de absorción de los electrones solvatados, con su pico de absorción característico en el infrarrojo, puede obtenerse con películas de cualquiera de estos metales, exceptuando el sodio. Los picos observados a longitudes de onda más cortas, peculiares de cada metal, se deben a la absorción de los aniones.

baja energía emitiendo un fotón. Las energías requeridas de los fotones suelen ser las correspondientes a las radiofrecuencias, desde dos megahertzios hasta 300 megahertzios. Cuando el campo magnético externo y la frecuencia de la radiación tienen precisamente los valores adecuados, se observa la resonancia y hay fuerte absorción de la radiación.

El experimentador que estudia la resonancia magnética nuclear en soluciones puede controlar el campo externo en el que el átomo se halla sumergido, pero este campo no es necesariamente el mismo que actúa sobre el núcleo. Los electrones que rodean el núcleo pueden oponerse al campo externo (comportamiento diamagnético) o pueden reforzarle (comportamiento paramagnético). Para compensar el apantallamiento diamagnético de los electrones, hay que reforzar el campo externo, mientras que la facilitación paramagnética requiere una disminución del mismo. Los consiguientes "desplazamientos químicos" suministran así información acerca del entorno inmediato del núcleo.

La técnica de la espectroscopía por resonancia magnética nuclear ha proporcionado una demostración convincente de la existencia de aniones de los metales alcalinos en solución. Vamos a ver cómo la resonancia magnética es capaz de distinguir los aniones verdaderos de las otras estructuras posibles, al considerar el estado de los electrones en un átomo de sodio en entornos distintos.

Para la resonancia magnética nuclear, un catión de sodio aislado puede considerarse como el estado de referencia de desplazamiento cero, ya que tiene la configuración electrónica del gas noble, que no es magnética. La situación es distinta para un catión de sodio en solución. Recordemos que los electrones más externos de este catión están en orbitales tipo *p* y tienen un momento angular orbital además de su momento angular de espín. En el catión aislado, los momentos angulares orbitales de los seis electrones tipo *p* se cancelan entre sí, pero la interacción de los electrones con las moléculas de disolvente altera este equilibrio. Como resultado de la interacción con el disolvente, los orbitales tipo *p* precesionan alrededor del núcleo creando un campo magnético que se suma al externo. En otras palabras, el catión en disolución presenta un desplazamiento químico paramagnético; lo mismo ocurre con los cationes confinados en una molécula de "cryptand". La magnitud del desplazamiento depende principalmente de las moléculas que estén en contacto directo con el catión.



LA ESTRUCTURA DE LOS CRISTALES obtenidos por precipitación de disoluciones de sodio puede considerarse constituida por capas alternativas de aniones sodio y de complejos formados por un "cryptand" y un catión sodio. En la mayoría de sus características fundamentales, estos cristales se comportan como las sales convencionales, tales como el cloruro sódico. Así, por ejemplo, deben tener un número igual de aniones y de complejos catión -y- "cryptand". Lo que distingue particularmente a este conjunto de cristales es la reactividad que presentan los aniones metálicos.

Un anión de metal alcalino tiene sólo orbitales completos; por ello, cuando está aislado debería tener un espectro de resonancia magnética nuclear con desplazamientos casi nulos respecto al espectro del catión aislado que se toma como referencia. Es más, el espectro debería continuar sin desplazamientos incluso cuando el anión se encuentra en solución. Los electrones de valencia del anión ocupan un orbital tipo *s* y no poseen momento angular orbital. Sus interacciones con las moléculas de disolvente no producen ningún campo magnético.

Los primeros estudios importantes de resonancia magnética nuclear de los cationes de sodio en "cryptand" fueron realizados en 1973 por Joseph M. Ceraso en mi laboratorio. Este investigador mostró que un mismo catión sodio permanece en la cavidad de una molécula de "cryptand" durante un tiempo suficientemente largo como para poder distinguirle de los demás iones sodio con situaciones distintas en la solución.

Con posterioridad, nuestra atención se dirigió a soluciones conteniendo, a la vez, cationes sodio atrapados en moléculas de "cryptand" y otras especies negativas que suponíamos habrían de ser aniones sodio.

Se observaron dos señales; una de ellas presentaba el desplazamiento paramagnético característico del ion Na^+ rodeado por el "cryptand". La otra señal se encontraba a una frecuencia no desplazada respecto a la calculada para el anión de sodio aislado. Sólo en un anión verdadero pueden los orbitales tipo *s* apantallar efectivamente al núcleo; otro tipo de estructuras podrían poseer también una carga negativa neta, pero deberían presentar los mismos rasgos magné-

ticos que los cationes en solución. Se han observado resultados similares con soluciones de rubidio y de cesio.

Si un metal alcalino en disolución puede comportarse como un anión, podría realizar el mismo papel en una sal, precipitada de esta solución. Podemos imaginar, por ejemplo, una sal hipotética, análoga al cloruro sódico, pero con sodios como ion positivo y como ion negativo. Los cálculos indican que esta sal sería menos estable que el sodio metálico que, por tanto, se formaría espontáneamente, pero la diferencia en estabilidad es muy ligera. Hemos preparado sales algo más complejas en las que el catión es un complejo de sodio y de un "cryptand". El "cryptand" aumenta lo suficiente la estabilidad de la sal como para evitar la formación del metal.

Mei Tak Lok y Frederick J. Tehan, en mi laboratorio, intentaron preparar una sal a partir de una solución apropiada de sodio y un "cryptand" en etilamina. Dado que sospechábamos que la sal tendería a descomponerse en metal y en "cryptand", evaporamos rápidamente el disolvente, esperando que de este modo el material se desecaría antes de que el catión pudiera escapar de la molécula de "cryptand". El resultado fue animador: obtuvimos una película de color dorado que tenía un aspecto muy distinto tanto del metal como del "cryptand".

De hecho, la preparación de esta sal resultó ser mucho más fácil de lo que esperábamos y el desecado rápido era innecesario. Ceraso, al preparar una solución de sodio y "cryptand" para una investigación de resonancia magnética nuclear, observó que se formaban espontáneamente precipitados de color oro cuando la solución se enfriaba hasta la temperatura del hielo seco (–78 grados Celsius). Lok y Tehan verificaron posteriormente que este material tiene la estructura de una sal cristalina de fórmula Na^+ 2,2,2-cryptand·Na^-.

Debido al brillo dorado de los cristales pensamos, al principio, que el nuevo compuesto se comportaría como un metal, un material que contiene una gran cantidad de electrones móviles. No obstante, se comprobó que era un semiconductor, en el que casi todos los electrones de valencia permanecen ligados, aunque muy débilmente, a sus respectivos átomos. En soluciones de etilamina, los cristales crecen en forma de plaqueta hexagonal. Tehan y B. L. Barnett han determinado la estructura del cristal, confirmando nuestra hipótesis de que el compuesto es una sal. Las capas de complejo catión-sodio-cryptand alternan con capas de aniones sodio. Los aniones sodio son parecidos a los iones ioduro en cuanto a su dimensión efectiva y a sus posiciones en la red cristalina.

Este nuevo compuesto de sodio sólido es muy reactivo frente al aire, la humedad y, en general, cualquier aceptor potencial de electrones. Incluso almacenado al vacío, no resulta indefinidamente estable más que a bajas temperaturas. A temperatura ambiente, los cristales oscurecen lentamente y pierden su brillo, debido a que los aniones sodio atacan las moléculas orgánicas de "cryptand". A pesar de ello, trabajando rápidamente hemos podido medir el punto de fusión de estos cristales. Es de 83 grados C., 15 grados por encima del punto de fusión de las moléculas de "cryptand" puro. La fusión va acompañada de la descomposición en sodio metálico y "cryptand" libre.

Michael G. DaGue, Michael R. Yemen y yo hemos encontrado explicación a la apariencia metálica de este compuesto no metálico. El brillo de los metales es una impresión subjetiva producida por un espectro de absorción óptica caracterizado por una absorción creciente escalonada desde las longitudes de onda más cortas hasta las más largas. En los metales, este perfil de absorción resulta de la absorción de la luz por los electrones libres que se hallan distribuidos a través de la red metálica.

Ocurre que los cristales de anión sodio tienen un espectro de absorción similar aunque no tengan electrones libres en abundancia. Por tanto, la apariencia metálica no proviene de una estructura metálica, sino de un mimetismo accidental. También contribuye a esta ilusión la estructura cristalina de la sal de sodio, que tiene grandes superficies cristalinas como las de un metal.

Tras la identificación inicial de la sal de sodio, hemos preparado compuestos que contienen potasio, rubidio y cesio que parecen tener una estructura análoga. La identificación definitiva de estos compuestos ha resultado más difícil dada su mayor inestabilidad. A la temperatura ambiente, la sal de sodio tiene una vida media, frente la descomposición, de algunos días; la sal de potasio correspondiente tiene una vida media de menos de un minuto. Las sales de los metales más pesados deben mantenerse constantemente a temperaturas por debajo de los –40 grados C., situación en la que todas ellas parecen ser estables durante períodos de tiempo muy largos.

Dado que las soluciones de aniones alcalinos y de sus cationes atrapados producen compuestos sólidos, parece posible obtener materiales cristalinos a partir de cationes atrapados y de electrones solvatados. Una posibilidad sería la obtención de algo parecido a un metal en el que los electrones libres, en vez de moverse entre los cationes metálicos de la red cristalina, se movieran a través de una red formada por complejos catión-"cryptand". Los intentos de obtener sólidos con esta estructura han fracasado hasta el presente. Por evaporación rápida del disolvente, hemos observado, sin embargo, películas residuales que parecen tener una estructura diferente. Estas películas son de un azul intenso y fuertemente paramagnéticas, lo que sugiere que los electrones están confinados en las proximidades de determinados cationes en vez de moverse libremente a través de la red. Los experimentos preliminares sugieren que estos sólidos contienen exactamente un electrón no apareado por cada catión; si esto se confirmase, las películas no serían metálicas en absoluto; se trataría del primer ejemplo de una nueva clase de sales: las sales de "electrones".

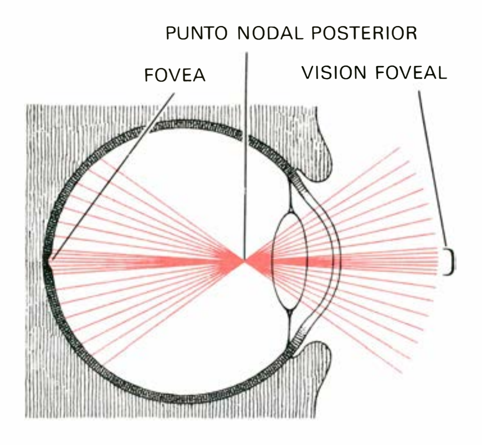
Actualmente se puede considerar demostrado que existen aniones de todos los metales alcalinos, excepto del litio. (Los cálculos indican que el anión litio debería ser también estable bajo las circunstancias adecuadas, por lo que nosotros continuamos buscando tal especie.) Es muy probable que, hasta el momento, no hayamos descubierto más que una fracción muy pequeña de los compuestos posibles que contienen aniones de metales alcalinos. Debido a que el tamaño de los "cryptand" puede alterarse fácilmente, probablemente podrían combinarse una variedad de cationes con cualquier anión. Una posibilidad obvia es la preparación de sales heterogéneas que contengan un catión atrapado de algún metal alcalino y un anión de otro. En términos más generales, los "cryptand" podrían confinar otros cationes distintos de los cationes de metales alcalinos. Por ejemplo, se podrían formar sales con tierras alcalinas tales como el bario, que tiene una carga +2, o con tierras raras, como el lantano, que tiene una carga +3. Las sales contendrían necesariamente una alta concentración de aniones de metales alcalinos: dos o tres por cada catión. Finalmente, podríamos intentar prescindir totalmente de los "cryptand". Ciertos cationes orgánicos, como, por ejemplo, los derivados orgánicos del ion amonio, deberían ser capaces de sobrevivir, sin otra protección, en presencia de aniones de metal alcalino o incluso en presencia de electrones.

El ojo compuesto de los insectos

Poseen centenares de facetas, cada una de las cuales representa un omatidio u ojo en miniatura. El tamaño, modelo y la orientación de los omatidios están determinados por las necesidades del sistema de vida del insecto

G. Adrian Horridge

Los insectos son legión: unos se muestran activos sólo con luz solar, otros en presencia y en ausencia de ella y, otros, por último, sólo durante el crepúsculo. Estos, como la mosca vulgar común, vuelan con gran rapidez y efectúan giros repentinos. Aquéllos, como las libélulas, vuelan alternativamente en una dirección, evolucionan y revolotean. Otros terceros, como la mantis religiosa, permanecen inmóviles durante horas. Todos esos insectos se apoyan en la visión para su supervivencia, y todos perciben el mundo a través de las múltiples facetas de sus ojos compuestos, aun cuando sus costumbres y sus necesidades visuales sean completamente diferentes. ¿Qué hace el ojo compuesto para trabajar como un mecanismo de percepción óptica? ¿Hasta qué punto da cuenta de las funciones por las cuales el insecto utiliza sus ojos? ¿Hasta qué punto la pequeñez de las facetas del ojo del insecto limita su sensibilidad? ¿Cuál es el límite para el trabajo del ojo del insecto, a bajas intensidades luminosas, con los pequeños cristalinos de las facetas? ¿Cómo llegan a un arreglo los ojos compuestos de los insectos entre la resolución óptica y la sensibilidad necesaria para superar el ruido óptico?



OJO DE UN ANIMAL VERTEBRADO. Posee una pequeña área en forma de copa en la retina en donde la visión es particularmente penetrante. En esa área, que es la de la fovea central, los receptores visuales forman paquetes apretados, por lo que su resolución, o facultad para distinguir finos detalles, es alta. Cada línea de color representa el eje óptico de un receptor determinado. El empaquetamiento de los receptores de la fóvea requiere que estos receptores sean estrechos, de suerte que la sensibilidad de cada receptor foveal sea menor que la de los receptores de cualquier otro lugar de la retina. Así pues, la fóvea de los vertebrados es un área especializada de la visión con alta resolución solamente en presencia de luz intensa.

El ojo compuesto está formado por omatidios: que son ojos diminutos que miran en diferentes direcciones. Cada omatidio consta de un cristalino que concentra la luz de varias células receptoras que comparten el órgano sensible a la luz común, conocido como rabdoma. En muchos insectos, cada omatidio es un elemento individual con su propio eje óptico; sólo se tratará aquí de los insectos dotados de este tipo de omatidio. La luz recogida por el rabdoma transmite una señal al cerebro del insecto a través de un nervio óptico. ¿Cómo podemos determinar las direcciones en las que está mirando un omatidio?

Si nos fijamos en el ojo de un insecto, vemos con frecuencia una mancha negra en el centro ocular. Aunque el insecto gire la cabeza, la mota negra siempre apunta en la dirección del observador. La mancha en cuestión recibe el nombre de pseudopupila: las facetas de ésta se ven negras porque reflejan menos luz en la dirección del observador que las otras facetas del resto del ojo. En otras palabras, el omatidio del centro de la pseudopupila está mirando directamente al observador.

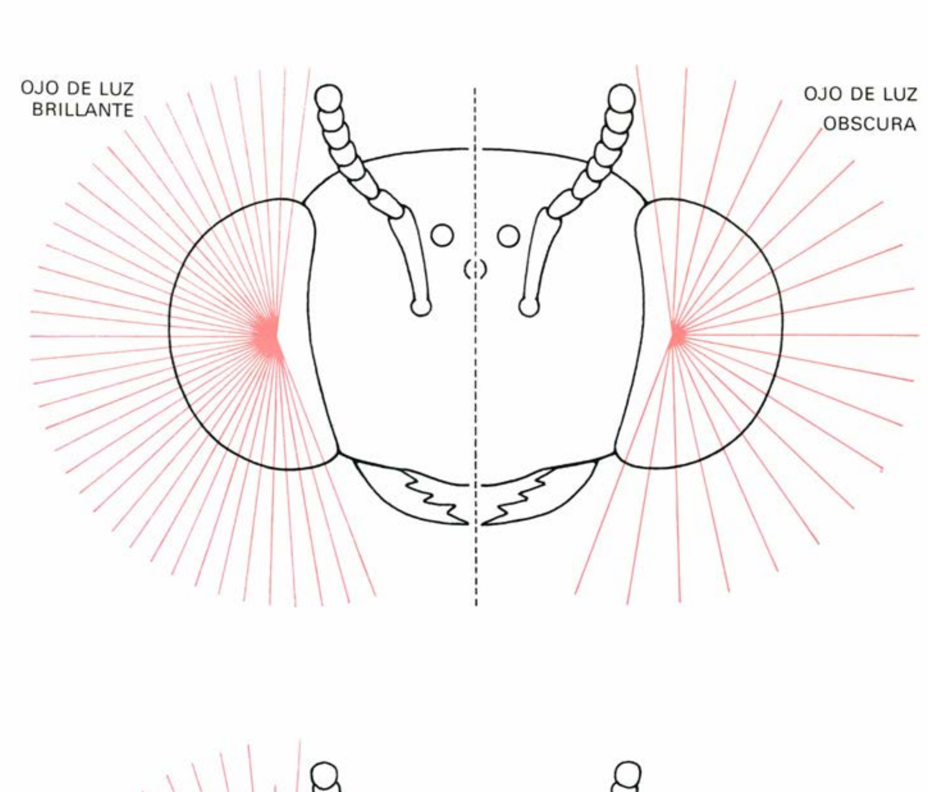
Para efectuar un esquema del eje óptico de cada omatidio, se coloca la cabeza del insecto (o todo el animal) en el centro de un goniómetro, que es un aparato para medir ángulos. Se mide el ángulo necesario para desplazar el centro de la pseudopupila la distancia comprendida entre cinco facetas; la medición se va repitiendo de cinco en cinco facetas hasta que se ha conseguido dibujar el mapa de todo el ojo. Los mismos datos se pueden recoger con más comodidad fotografiando el ojo a intervalos de 5 o 10 grados. En este caso, se marcan las facetas con partículas de polvo, y se miden, a partir de las fotografías, las posiciones de la pseudopupila con respecto al modelo de las facetas marcadas. La posición del eje de cada quinta faceta se dibuja después sobre una parrilla de coordenadas angulares. El modelo resultante es un mapa del camino seguido por el ojo compuesto para recorrer el campo visual. Hasta ahora sólo se han hecho unos cuantos mapas parciales, como el preparado para la mosca común por los biofísicos holandeses D. G. M. Beersma, Doekele G. Stavenga y Jan W. Kuiper. Sólo se pueden efectuar mapas de bandas estrechas, sin llegar a afrontar los problemas normales con que se topa uno cuando traslada el mapa de una esfera sobre una superficie plana.

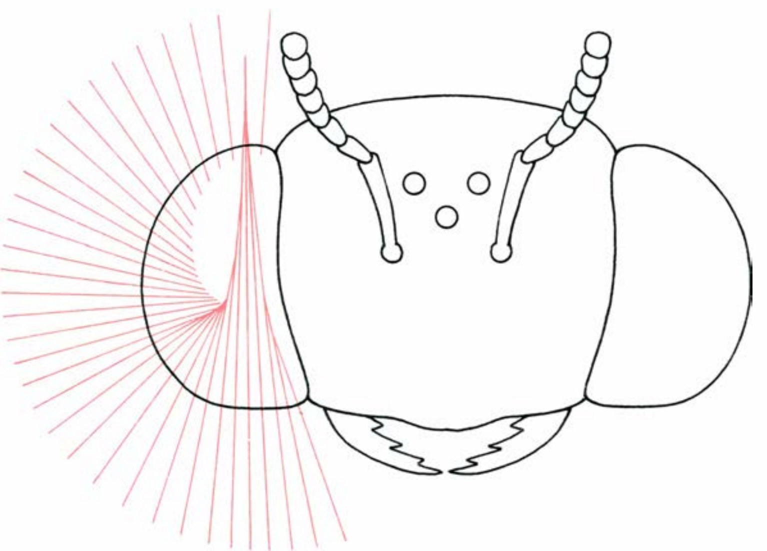
Para comprender el funcionamiento del ojo de un insecto son de la máxima importancia dos ángulos. El primer ángulo es el que forman los ejes ópticos de dos omatidios adyacentes; se trata del llamado ángulo interomatidial, que se simboliza $\Delta\Phi$; determina la densidad con la que el ojo compuesto se da cuenta del campo visual. El segundo ángulo define el campo visual de un solo omatidio, y se simboliza $\Delta\rho$. La forma del campo de la sensibilidad del omatidio para una fuente luminosa puntual como una función del ángulo del eje óptico. La sensibilidad del omatidio encuentra su intensidad máxima a lo largo del eje óptico, y decrece con el ángulo que forma con el eje. La amplitud del campo está definida por el ángulo subtendido por el diámetro que atraviesa el campo en donde la sensibilidad ha decaído en un 50 por ciento.

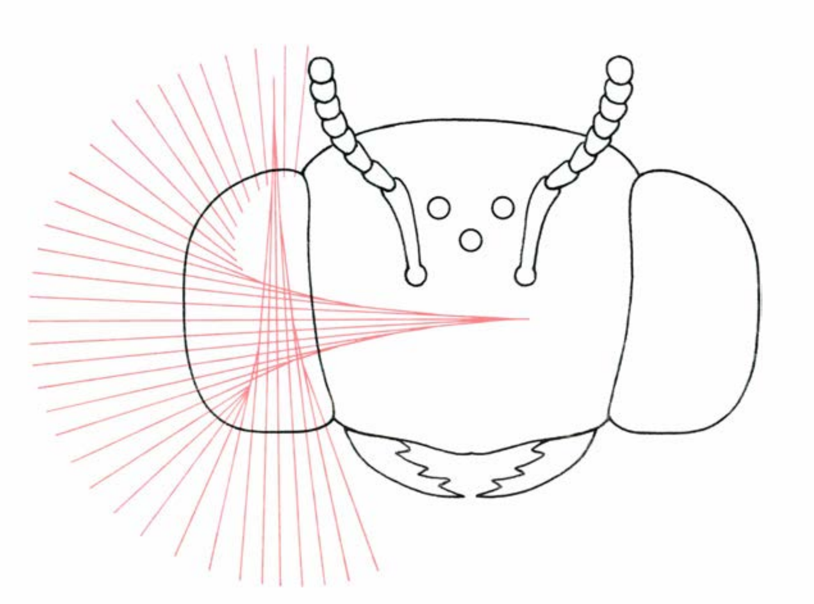
Los mapas del ojo sorprenden por el sistema que sigue el insecto para percibir el mundo. Ante todo, los mapas indican que el modelo de los ejes ópticos no coincide con el modelo de las facetas hexagonales de la superficie ocular. Las facetas, que están separadas a igual distancia en la superficie del ojo, no poseen necesariamente ejes ópticos con el mismo ángulo interomatidial. De hecho, en una región del ojo los ejes ópticos del omatidio pueden estar inclinados el uno hacia el otro, de suerte que resulten casi paralelos; por consiguiente, ofrecen un segmento del mundo visual y no la parte distributiva del enfoque. En esa región, la pseudopupila suele ser mayor que en cualquier otra parte del ojo, porque hay más facetas que miran en esa dirección.

Dicha área del ojo compuesto de los insectos tiene la misma función que la fóvea central de la retina de los vertebrados (incluido el hombre). En los vertebrados, la fóvea es un área en donde los receptores visuales están formando paquetes apretados para favorecer una visión aguda. En los ojos de los animales diurnos (incluido el hombre), los receptores deben ser muy estrechos para que puedan unirse compactamente los unos a los otros; de ahí que la sensibilidad de cada receptor sea reducida. Y ello explica que la fóvea de los vertebrados sea una adaptación para la visión fina en presencia de una luz esplendente.

Sin embargo, en los insectos, las facetas de la fóvea suelen ser mayores que las del resto del ojo. Al reducirse el ángulo





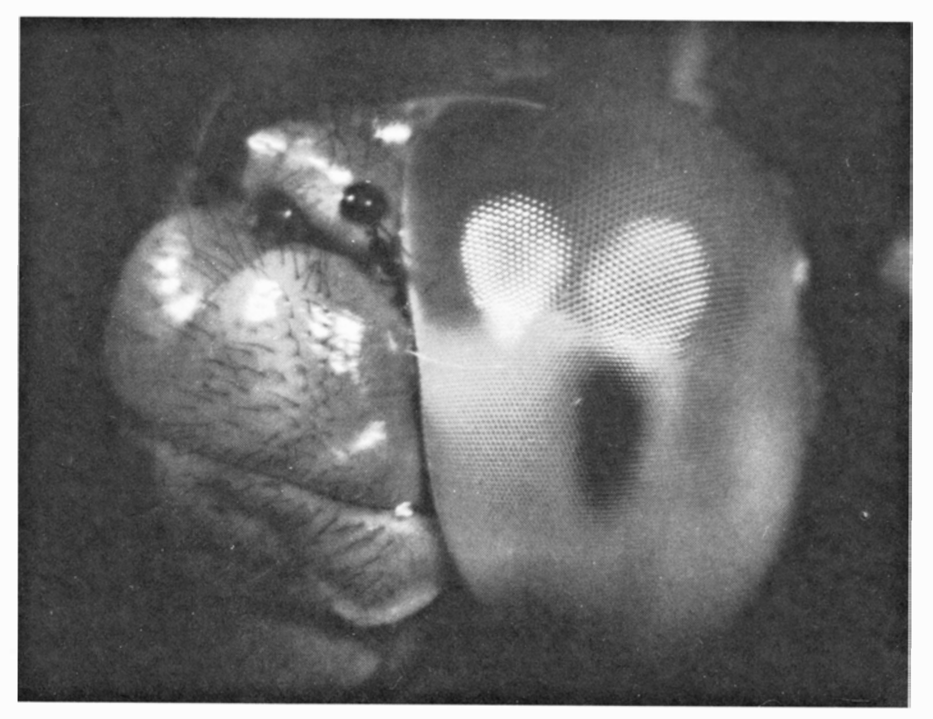


EL OJO COMPUESTO de un insecto está formado por omatidios. El tamaño real de cada parte de dicho ojo queda determinado por su radio óptico, que es el radio medido desde el punto en que corta el eje óptico de un omatidio adyacente. Se pueden distinguir cuatro tipos de ojos compuestos, de adelante atrás. Si el ojo del insecto fuera hemisférico y los omatidios de igual apertura miraran perpendicularmente a la superficie del ojo, todo el ojo tendría un radio óptico, y podría observar el mundo visual con igual densidad en todas las direcciones *(arriba)*. La sensibilidad de un ojo compuesto depende en parte del tamaño de sus facetas y, por consiguiente, del tamaño global del mismo para un determinado número de facetas. Sin embargo, las mediciones efectuadas en tales ojos compuestos indican que, en muchos de ellos, el radio óptico de la parte frontal del ojo es mayor que en el resto del ojo *(centro)*. Por esta razón, el campo de visión abarcado por la parte frontal del ojo se aprecia más densamente que la que observan las demás regiones del ojo. Este tipo de área recibe el nombre de fóvea por analogía con la fóvea del ojo de los vertebrados. En la fóvea de la mayoría de ojos de los insectos, las facetas son mayores, lo que es posible al poseer los ojos un radio óptico mayor. Algunos insectos *(abajo)* tienen una segunda fóvea que mira hacia arriba o hacia un lado; las grandes libélulas poseen tres tipos de fóveas.

interomatidial de la fóvea, la región del ojo que ocupa la fóvea es más plana que el resto del ojo; en otras palabras, posee un radio de curvatura mayor que el del resto del ojo. Por tanto, para que pueda alojarse la fóvea en una zona del ojo compuesto, éste debe tener un menor radio óptico de curvatura en cualquier otro lugar. De ello resulta que unas zonas del ojo quedan empobrecidas en puntos de recepción de impresiones en aras de un enriquecimiento por parte de otras.

Las facetas mayores posibilitan que sus omatidios dispongan de un mayor poder de resolución. Para los cristalinos, la capacidad de resolver finos detalles depende de dos factores: la apertura o diámetro del cristalino *(D)* y la longitud de onda (λ) de la luz que reciben. (En razón de la dificultad para medir la apertura de una lente hexagonal, *D* se toma como la distancia entre dos centros de facetas adyacentes.) La luz procedente de un punto lejano y que pasa a través de la lente no se concentra en un punto sino que mancha de modo borroso un círculo en el plano focal de la lente. La intensidad de la luz en este círculo borroso (o borrón) posee su máximo en el centro y remite hacia los bordes. Hay que definir la amplitud de este círculo borroso como el diámetro en donde la intensidad de la luz ha caído al 50 por ciento. Esta anchura recibe la denominación de $\Delta\ \alpha$. Por teoría de la difracción se sabe que la amplitud angular del círculo borrón, en radianes, es igual a la longitud de onda de la luz dividida por el diámetro de la lente ($\Delta\ \alpha = \lambda / D$). (Un radián es una circunferencia dividida por π, 57 grados, aproximadamente.) De aquí que para ciertas longitudes de onda de la luz, cuanto mayor sea la lente menor será el círculo borroso en cada punto luminoso y más perfilada la imagen.





LA PSEUDOPUPILA DEL OJO DE UN INSECTO aparece como una zona oscura en el ojo del odonato (libélula) *Diplacodes nematodes.* Conforme el insecto va doblando la cabeza, la pseudopupila se va desplazando de forma que siempre quede orientada hacia el observador. La pseudopupila es negra porque las facetas de esta zona reflejan menos luz en la dirección del observador que las facetas del resto del ojo. Las facetas del centro de la pseudopupila están apuntando directamente al observador. En general, donde la pseudopupila es mayor *(arriba)*, mayor número de omatidios mira en esa dirección. Al lado del ojo *(abajo)*, la pseudopupila está prolongada verticalmente debido a que los ejes ópticos de las filas verticales forman entre sí un ángulo que es la mitad del que se encuentra entre los ejes visuales de los ejes de los omatidios de las filas horizontales.

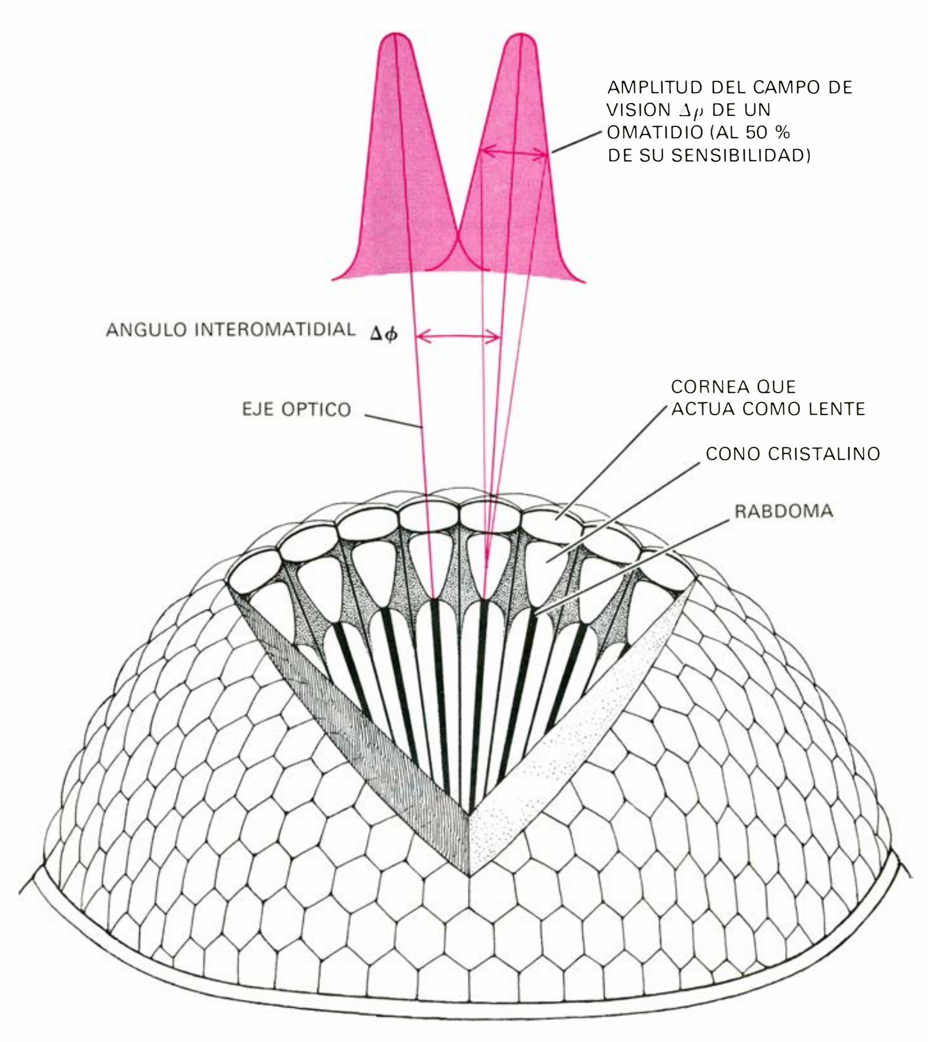
Resulta muy claro que, en los insectos que confían en la visión aguda, la realización óptica de la lente de cada omatidio es excelente; además, la estructura de recepción que toma la luz (que está en efecto en el extremo de un rabdoma) radica en el plano focal del cristalino. La resolución de cada omatidio se ve doblemente limitada. En primer lugar, por el diámetro del círculo borroso: una gran faceta del ojo de un insecto puede tener un poder de resolución mayor que otra faceta más pequeña, ya que concentra la luz en una mancha más pequeña. En segundo lugar, la limitación tiene que ver con cuán pequeño pueda ser el receptor hasta perder su capacidad de funcionar como conductor de la luz; el diámetro mínimo del receptor es de unos 0,5 micrómetros.

Si un punto luminoso se encuentra en el eje óptico del omatidio, la parte más intensa, o más brillante, del círculo borroso estará centrada en la parte superior del rabdoma. Conforme el punto luminoso se va moviendo a través del campo de visión, se va apartando del eje óptico y el máximo del círculo borroso se aleja del rabdoma. Así, el rabdoma interceptará cada vez un segmento menor y menos intenso del círculo borroso. Si el rabdoma es relativamente estrecho, puede medir con precisión la variación de la intensidad de la luz a través del círculo y su resolución está limitada solamente por la difracción de la luz que dio origen al círculo.

De la misma manera, un rabdoma estrecho percibirá sólo una pequeña fracción de la luz total del círculo, y su sensibilidad será baja. Si el rabdoma es más ancho, abarcará más luz del círculo, pero su resolución será más baja, debido a que no podrá medir los detalles con precisión. De aquí que la sensibilidad angular óptima del rabdoma sea una función de la amplitud, $\Delta \alpha$, del círculo, y del diámetro, *d*, del rabdoma. En cada ojo existe un compromiso entre la resolución, que está limitada sólo por la difracción o por la apertura mínima del rabdoma, y la sensibilidad que, como veremos, es muy importante en ojos con cristalinos pequeños.

Este compromiso o acuerdo puede describirse gráficamente con una comparación, verbigracia: las formas que hay de registrar el perfil del chorro del rociador de una ducha mediante recipientes de diferentes tamaños. Sea el diámetro del chorro equivalente a la anchura del círculo borroso y, las gotas de agua, equivalentes a los fotones de luz. Si cortamos el chorro con una pequeña taza, que lo haga rápidamente, se podrá medir, o resolver, la intensidad del perfil del chorro de la ducha, y distinguirá los límites de varios chorros que estén muy juntos. La taza llegará a interceptar, en cualquier punto, una sola porción de la ducha, es decir, la sensibilidad de la misma será baja. El chorro se coge con mucho mayor eficacia con un tubo ancho. El tubo es así más "sensible" que la taza; ahora bien, por coger tanta agua no puede distinguir detalles o separar los chorrillos que están muy juntos. No obstante, cuando el flujo es bajo sólo se pueden detectar los chorros grandes, mostrándose el tubo como el medio más adecuado para medir la distribución de los chorros irrigantes.

Si el diámetro *d* de un rabdoma circular es más estrecho comparado con la anchura del círculo borroso $\Delta \alpha$, el campo de visión del rabdoma $\Delta\rho$ será igual a la amplitud angular del círculo borroso. Si la extensión del rabdoma es lo suficientemente ancha como para abarcar el 95 por ciento del total de la luz del círculo, el campo de visión del rabdoma $R\rho$ resultará igual a dos veces la amplitud angular del círculo borroso en el 50 por ciento de intensidad, es decir, 2 $\Delta\alpha$. Si se aumenta el diámetro del rabdoma más allá de este valor, se obtiene una pérdida progresiva de poder resolutivo sin que ello importe el correspondiente aumento de sensibilidad para un punto de luz concreto. En la analogía de la ducha, si tuviéramos un tubo mayor que el diámetro del chorro no aumentaría la cantidad de agua que el tubo puede coger

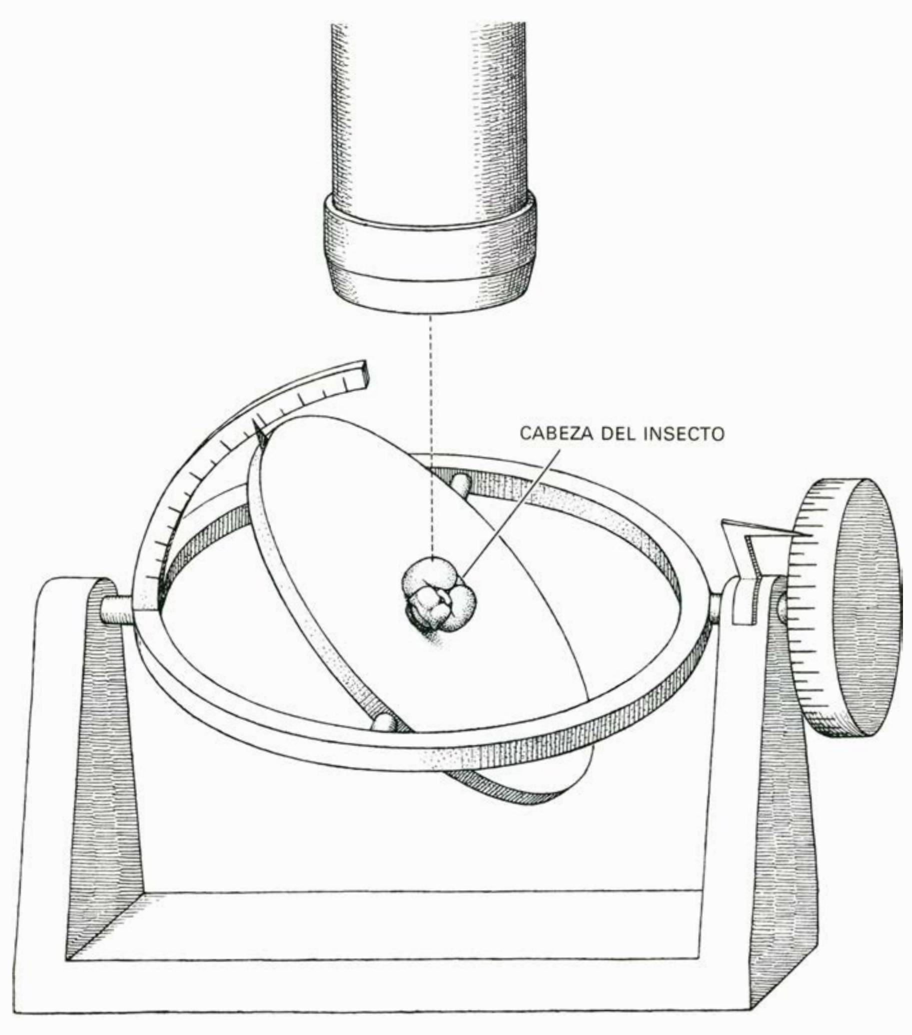


LA ESTRUCTURA DEL OJO DE UN INSECTO determina con cuánto detalle puede ver éste. Cada omatidio posee una córnea que también sirve como lente para enfocar la luz a través de un cono transparente hacia un elemento sensible a la luz: el rabdoma. El eje óptico del omatidio es la línea que pasa por el centro de la lente hasta el rabdoma. El ángulo interomatidial, $\Delta \Phi$, es el ángulo que se forma entre el eje óptico de dos omatidios adyacentes. Las curvas de color, por encima de cada omatidio, indican cómo decrece la sensibilidad del omatidio con la distancia angular desde el eje óptico del omatidio. El campo visual de cada uno de los omatidios está definido como el ángulo que engloba el campo donde la sensibilidad ha descendido hasta el 50 por ciento.

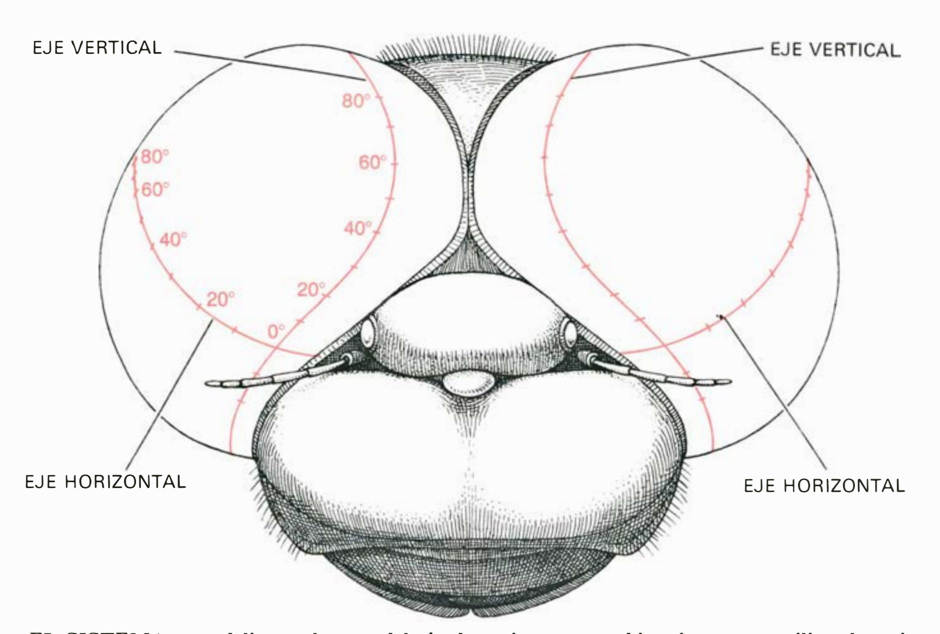
de la ducha. Sin embargo, hacer del rabdoma un "tubo" mayor implica planear una buena estratagema para ver al menos algo cuando los fotones andan tan escasos que sólo se pueden apreciar los haces mayores.

Hasta aquí, en este examen del origen del campo de visión del insecto a partir del círculo borroso, ha sido necesario eludir los problemas que presenta la refracción, o desviación, de la luz cuando pasa del aire al ojo del insecto. El ángulo de apertura del círculo borroso $\Delta \alpha$ (dentro del ojo) y el ángulo del campo visual (fuera del ojo) se miden desde el *punto nodal posterior,* el cual, en una lente unilateral como ésta, es un punto muy cercano al centro de curvatura de la lente. Los rayos luminosos que pasan por el punto nodal posterior, dentro del ojo, son paralelos a los mismos rayos fuera del ojo: por esta razón, el punto nodal posterior es muy útil para relacionar los ángulos medidos fuera del ojo con los ángulos medidos en el interior del mismo. El diámetro de un receptor ancho, limitado por el punto nodal posterior, es el tamaño del campo visual $\Delta\rho$.

Supongamos que en un omatidio típico de alto poder resolutivo el cristalino tiene 28,5 micrómetros de apertura. Si la longitud de onda es de 0,5 micrómetros (la longitud de onda de la luz verde), el diámetro del círculo borroso $\Delta \alpha$ será de un grado. Si, en este omatidio, el diámetro del rabdoma (delimitado por el punto nodal posterior) es igual a la luz del círculo borroso al 50 por ciento de la intensidad $\Delta\alpha$, entonces, y según la fórmula $\Delta\rho^2 = \Delta\alpha^2 + (d/f)^2$, su campo visual sería igual a 1,4 grados. Este rabdoma tendría un diámetro de un micrómetro y una distancia focal *f* (medida desde el punto nodal posterior al extremo del



LOS MAPAS DE LOS OJOS DE LOS INSECTOS se realizan montando la cabeza del insecto en la plataforma de un goniómetro (brújula calibrada) y fotografiándolo desde una serie de ángulos. El ángulo requerido para mover el centro de la pseudopupila cinco facetas determina el ángulo interomatidial entre uno de cada cinco omatidios. Una vez realizado ello, la posición del eje óptico de cada quinto omatidio se transcribe luego fielmente sobre un mapa preparado al efecto.



EL SISTEMA para delinear el mapa del ojo de un insecto es arbitrario, pero se utilizan los principales ejes funcionales del ojo. El cero del eje vertical coincide normalmente con el cero del eje horizontal en la posición del insecto en la que la fóvea mira hacia adelante. El eje vertical cero en los mapas en coordenadas angulares es la línea de facetas del plano en el que se cortan las direcciones hacia adelante y vertical. No es una línea recta bajo el ojo. El eje horizontal cero en los mapas es la línea que une las facetas que miran horizontalmente. No es una fila horizontal de facetas a través del ojo, pues el eje óptico puede quedar inclinado con respecto a la superficie del ojo.

rabdoma) de 57 micrómetros. Un rabdoma con un diámetro inferior a un micrómetro no sirve grandemente como guía luminoso, de ahí que tales rabdomas angostos apenas sí se encuentran en los ojos de los insectos. Una mayor apertura en la faceta generaría un círculo borroso más pequeño, exigiendo a su vez un rabdoma más estrecho. Por eso, si la difracción de la luz limita el trabajo del omatidio, por lógica, habría que descartar la existencia de insectos con omatidios cuyas facetas superaran los 30 micrómetros, porque los rabdomas eficaces o activos no pueden estrecharse tanto como para llegar a emparejarse con las facetas.

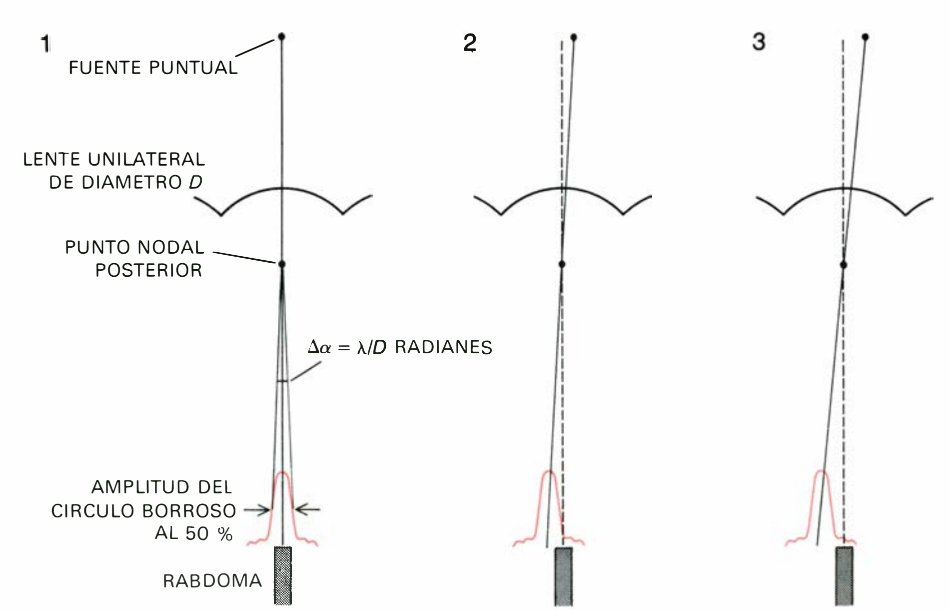
Pero la realidad lo desmiente, porque suelen hallarse fácilmente facetas de 80 micrómetros de diámetro, especialmente en la fóvea. Algunos organismos, distintos de los insectos, como, por ejemplo, el estomatópodo *Odontodactylus* y el anfípodo marino de las profundidades *Phronima*, poseen un ojo compuesto con facetas de más de 120 micrómetros. Todos los ojos compuestos tienen menos facetas, y mayores, que lo que tendría que ser según la teoría de la difracción. ¿Por qué se encuentran estas aperturas mayores? Porque un aumento ulterior del diámetro de la lente incrementa la sensibilidad del omatidio con relación a la superficie de la lente (D^2). La consideración de los puntos luminosos y el del límite de difracción no han de tomarse en todo rigor a la hora de comprender el funcionamiento del ojo compuesto. Los distintos omatidios se adaptan, según parece, de múltiples formas para conseguir captar hasta el último fotón de luz. En la comparación de la ducha, debemos volver desde el proceso de recoger el chorro de una única ducha pequeña hasta distinguir entre duchas de escaso flujo, de cualquier tamaño y en diferentes combinaciones. Cuando el chorro (fotones) es tan pequeño que sólo se pueden detectar las duchas de mayor flujo, importa entonces la sensibilidad.

Hasta aquí me he concentrado en las necesidades ópticas y en las características de un solo omatidio. Pero el ojo compuesto consta de centenares de omatidios, que miran de modo más o menos igual en muchas direcciones. La ventaja de esa disposición, es que opera por un proceso paralelo: divide la información que proviene del mundo exterior en muchos canales paralelos, analiza la información de cada canal y recombina parte de la misma en el cerebro. Por consiguiente, un insecto puede conocer todo su mundo visual en cada instante, mien-

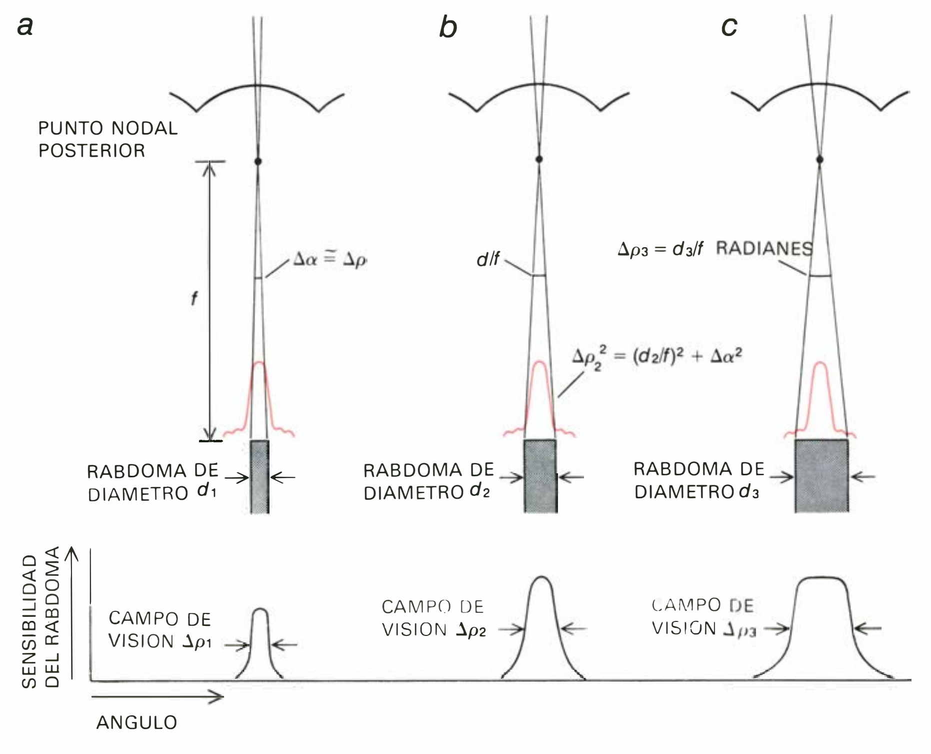
tras que en los vertebrados (entre los que se cuentan los humanos) sólo se puede obtener esta información a base de mover continuamente ojos y cabeza en una serie de exploraciones.

¿Qué es lo que rige el espacio y distribución de los omatidios en la superficie ocular? En 1894, A. Mallock (y, posteriormente, Horace B. Barlow, Hugo de Vries, Jan Kuiper y muchos otros) propusieron que la distribución angular de los omatidios podría calcularse a partir de los límites impuestos por la difracción, por lo que la distribución angular de lentes adyacentes ($\Delta\Phi$) no tiene que tener más de dos veces el espacio ($\Delta\theta$) del dibujo más preciso que cada uno pueda resolver. El razonamiento fue básicamente el siguiente:

El mundo que el insecto percibe puede figurarse como un modelo complejo de contrastes entre luz y oscuridad. Modelo que puede verse, aproximadamente, como la suma de cierto número de dibujos de rayas normales, cada uno con una distribución sinusoidal de intensidad entre luz y oscuridad y un período normal $\Delta\,\theta$. En una disposición de receptores para ver un modelo rayado y reconstruirlo en bandas por un procedimiento paralelo simultáneo, la separación angular mínima entre receptores (es decir, el ángulo interomatidial $\Delta\ \Phi$) debe ser igual al ángulo que comprende la mitad de la separación del dibujo listado (eso es, $1/2\ \Delta\,\theta$). Con esta mínima separación, dos picos (bandas luminosas) en la gráfica de la intensidad luminosa de las rayas, junto con las que están entre las dos primeras bandas (una raya oscura), se dirige a tres receptores adyacentes. Si el ángulo interomatidial fuese más pequeño, habrían receptores más que suficientes para resolver el dibujo; si el ángulo interomatidial fuese mayor, habría pocos receptores y no podrían resolverlo. Esta idea se conoce como la teoría de Whittaker-Shannon, y en dos dimensiones lo aplica a receptores dispuestos en modelo cuadrado. Tal como ha sido demostrado recientemente por Allan Snyder, de la Australian National University, y William H. Miller, de la Yale University School of Medicine, para que un modelo de receptores hexagonal alcance la misma acuidad visual que un modelo cuadrado, el hexagonal precisa más receptores, en un factor de 2 dividido por la raíz cuadrada de 3. Los mapas de ojos de insectos indican que el modelo de los ejes ópticos de los omatidios radica entre los sistemas cuadrado y hexagonal. El modelo hexagonal se retuerce en otro de cuadrados oblicuos, especialmente en torno al ecuador del ojo.



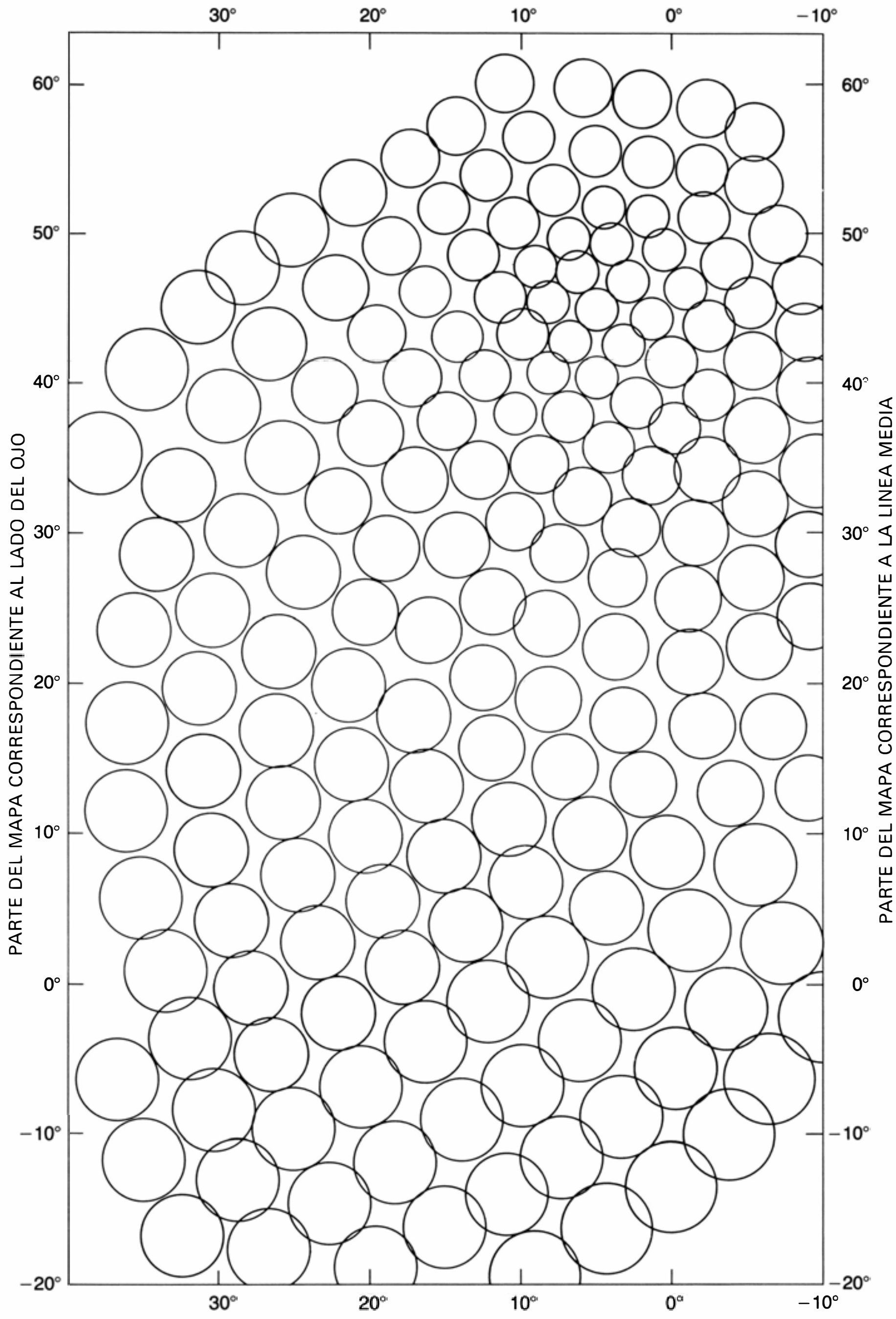
LA LUZ SE DIFRACTA al pasar a través de la lente o cristalino de un omatidio. En consecuencia, la luz proveniente de un punto luminoso no se registra como un punto sino como un modelo circular, el círculo borrón. La intensidad de este círculo borrón *(curvas en color)* es fuerte en el centro y así remite hacia los bordes. La anchura del círculo borrón ($\Delta\alpha$) se define como su diámetro en donde la intensidad ha caído al 50 por ciento. Esta anchura angular (en radianes) es igual a la longitud de onda de la luz del punto luminoso dividida por el diámetro D de la lente. Cuando el círculo borrón se encuentra en el eje óptico *(1)*, el rabdoma recibe la mayor cantidad de luz. Si el punto luminoso se mueve hacia el eje óptico *(2 y 3)*, el círculo borrón se desplaza desde el extremo del rabdoma y la cantidad de luz que recibe el rabdoma disminuye. Los rayos son dibujados a través del punto nodal posterior, que se halla cerca del centro de curvatura de la lente. Los rayos que pasan por este punto recorren, en el interior del ojo, un camino paralelo al seguido en el exterior.



LA SENSIBILIDAD Y RESOLUCION DEL RABDOMA son función de su diámetro y la distancia del punto nodal posterior. Un rabdoma estrecho *(a)* con un diámetro d_1 inferior a la anchura del círculo borroso, al 50 por ciento de intensidad, recibirá sólo una parte de la luz en el círculo borroso pero no responderá bien a la variación de luz a través del círculo borrón y, por tanto, tendrá alta resolución, pero sensibilidad baja. Un omatidio con tal rabdoma se halla cerca del límite de difracción. Un rabdoma más ancho *(b)*, con un diámetro d_2 igual a dos veces la anchura del círculo borroso, recibirá el 95 por ciento de la luz del círculo, pero tendrá un campo ligeramente más ancho, y, por consiguiente, un poder de resolución más bajo. Un omatidio *(c)*, con un rabdoma de diámetro mayor d_3, tendrá un campo de visión ancho que no resulta mejor que el *b* para las fuentes luminosas puntuales, pero que es mejor para la detección de objetos que se encuentran con luz apagada. La sensibilidad angular, o la forma del campo, de cada rabdoma (medido a lo largo de la apertura al 50 por ciento de sensibilidad) se muestra en las curvas de la parte inferior de la ilustración. La sensibilidad para objetos grandes aumenta con el tamaño del rabdoma, según la relación $(d/f)^2$, de la misma manera que la película fotográfica es más sensible cuando el grano resulta mayor.

Por teoría de la difracción se conoce también que el espaciado de las rayas $\Delta\,\theta$ en el modelo más fino que se puede resolver a cualquier longitud de onda de la luz con lentes de apertura D es igual a la longitud de onda de la luz λ dividida por la apertura de la lente ($\Delta\,\theta = \lambda/D$). Ya que el ángulo interomatidial mínimo $\Delta\Phi$ debe ser igual a la mitad de la separación entre rayas, luego, en orden al límite de difracción del ojo para poder ver, el producto de la apertura de la lente y el ángulo interomatidial deben ser iguales a la mitad de la longitud de onda de la luz (es decir $D\,\Delta\Phi = \lambda/2$). Los investigadores pensaron que éste era el caso.

Si sabemos, no obstante, que el ojo debe operar a un más alto nivel de modulación de la luz para vencer el ruido de los fotones, podemos decir que $D\,\Delta\Phi$ es igual a λ, de suerte que $D\,\Delta\Phi$ es ahora igual a 0,5 micrómetros. El producto $D\,\Delta\Phi$ sería constante por esta razón para un ojo que funcione con un nivel de modulación dado, y sería 0,25 micrómetros para los modelos cuadrados de ejes trabajando al límite de difracción. Ambos, D y $\Delta\,\Phi$, se pueden medir desde fuera del ojo. En términos geométricos, el ángulo interomatidial en radianes es el diámetro de la faceta D dividido por el radio óptico R. Por ello D^2/R podría ser constante para ojos de diferentes tamaños que trabajaran a la misma intensidad. En otras palabras, conociendo la apertura de la lente del omatidio D y el ángulo interomatidial $\Delta\,\Phi$, se puede determinar gran parte de la estructura del ojo, a partir del producto de los dos. Esto se aplica, al propio tiempo, para insectos distintos y para distintas regiones del ojo de un mismo insecto.

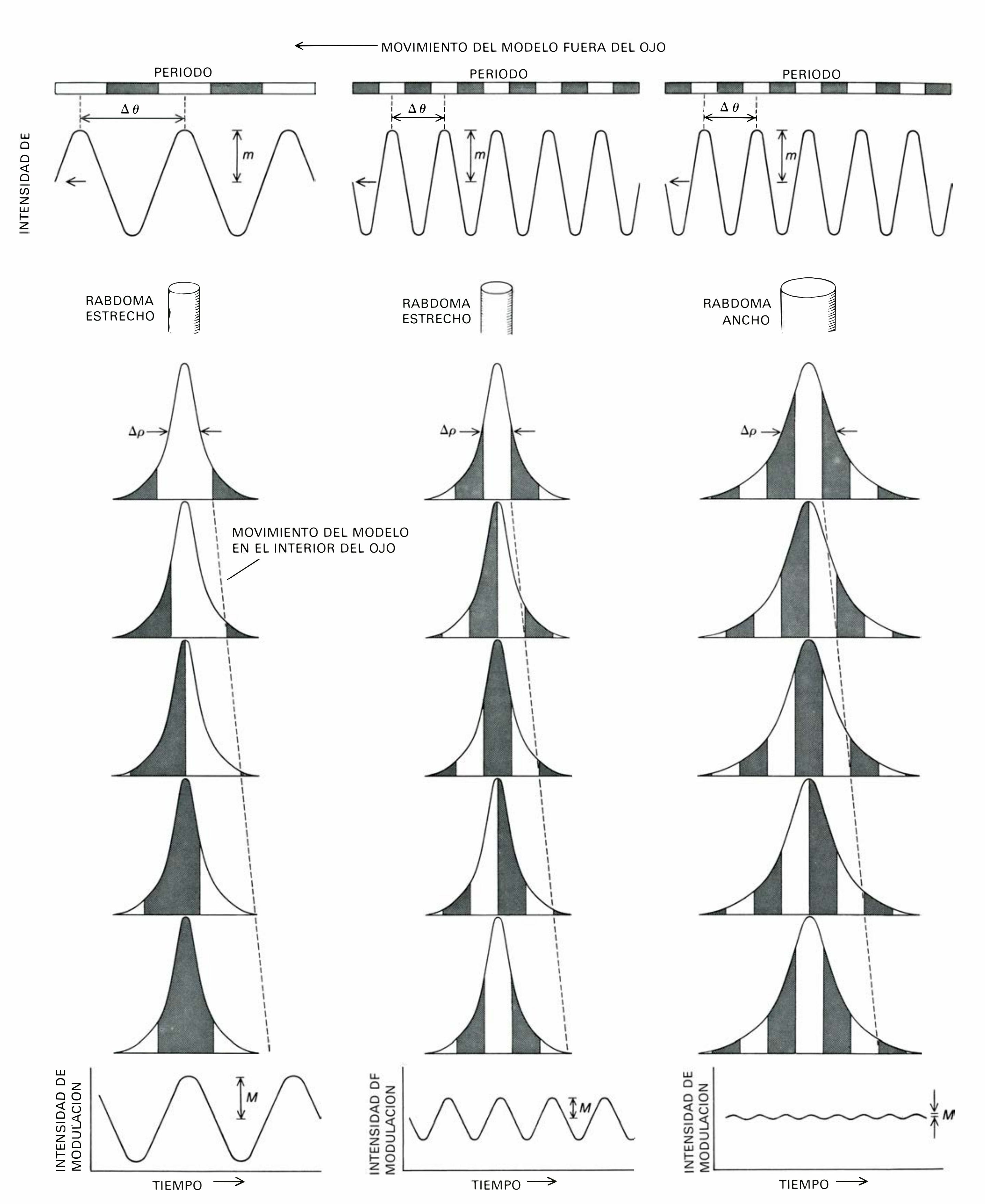


EL MAPA DEL OJO muestra la parte frontal del ojo de la libélula *Austrogomphus* que posee una fóvea que mira oblicuamente hacia arriba y hacia adelante. Cada círculo representa el campo más estrecho posible de uno de cada cinco omatidios. Para compensar el factor cinco, el diámetro es igual a cinco veces la longitud de onda de la luz verde (que se toma como 0,5 micrómetros) dividido por el diámetro de la lente del omatidio, 0,5 α/D. Los círculos más pequeños significan que las facetas de este área del ojo del insecto son mayores y poseen un poder de resolución más alto que las de áreas con círculos mayores. El centro de cada círculo representa la posición del eje óptico de cada quinto omatidio; la distancia desde el centro de cada círculo al centro del siguiente es, por tanto, igual a cinco veces el ángulo interomatidial, o 5 $\Delta\Phi$. El total que separa o superpone los círculos da el valor del parámetro $D\Delta\Phi$, que es una estimación crítica de cuánto se aparta el ojo del insecto del límite de difracción en esta zona. En este ojo sólo hay alrededor de una cuarta parte de facetas que quedarían hacia el límite de difracción. El centro de la fóvea está en el centro de una agrupación de pequeños círculos que se hallan próximos a 0 grados, 50 grados de los ejes en espacio angular.

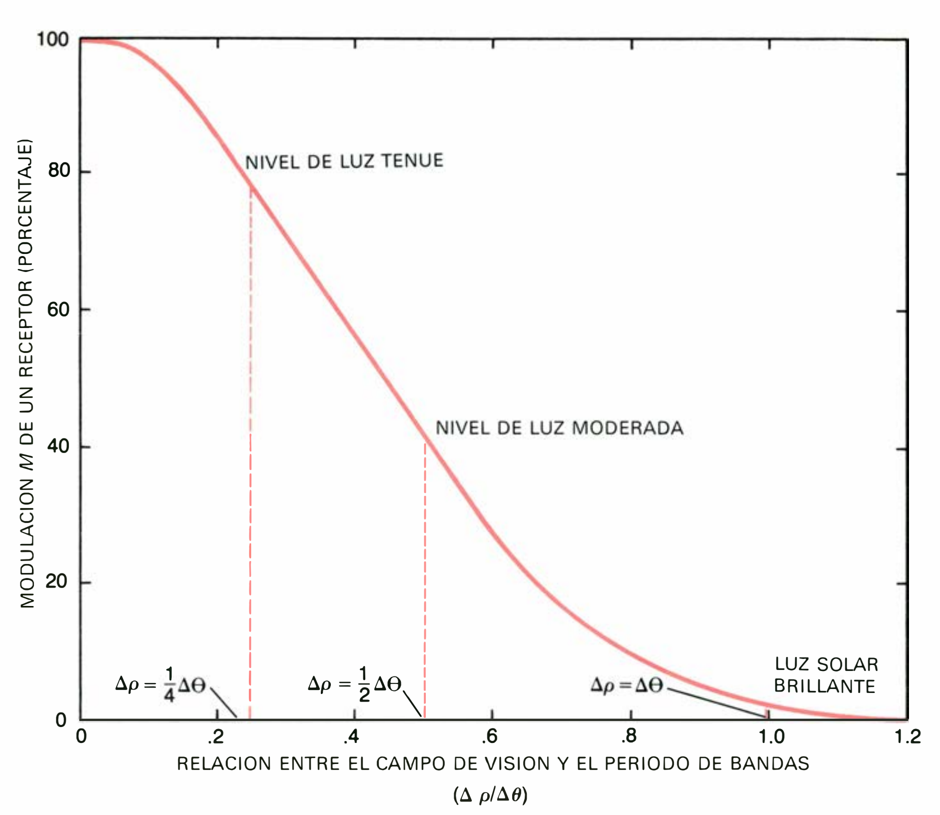
Si dibujásemos la estructura del ojo compuesto de un insecto, para reconstruirla detalladamente en el límite de difracción de una longitud de onda de 0,5 micrómetros, el parámetro del ojo $D\,\Delta\,\Phi$ quedaría fijado en un valor entre 0,25 y 0,31 micrómetros, según se distribuyese el modelo del eje óptico entre cuadrados y hexágonos. Los mapas del ojo revelan, sin embargo, que no existen insectos con ojos compuestos cuya estructura quede fijada en estos límites. La luz solar no es nunca lo suficientemente intensa como para eliminar el ruido de los fotones en un factor. Sólo las fóveas de algunos insectos activos en luz solar se acercan al límite de difracción.

Cuando un modelo de listas con un espaciado $\Delta\,\theta$ atraviesa el campo de visión de un insecto, el rabdoma percibe el dibujo como una modulación, u oscilación, que ha cubierto una intensidad luminosa media. Si el espaciado del modelo listado $\Delta\,\theta$ es grande con relación a la anchura del campo $\Delta\rho$, la amplitud de la modulación desde la luz a la oscuridad será también grande. Si el espaciado del modelo listado es pequeño, la amplitud de la modulación será pequeña. La rapidez de movimiento no es importante mientras se alcancen niveles en que el receptor responda a tiempo.

La intensidad lumínica media depende de la iluminación ambiental de las rayas y de la anchura del campo. La amplitud de la modulación M, que experimenta el receptor es, de hecho, la fluctuación de la intensidad de la luz a partir de una intensidad media. En concreto, la modulación relativa que experimenta el re-



MODELO DE BANDAS CLARAS Y OSCURAS *(arriba)*. **El rabdoma detecta el movimiento a través del campo de visión del omatidio, como una modulación de la intensidad de la luz a través del tiempo. Las bandas tienen una distribución sinusoidal de intensidad de amplitud** *m*, **desde la luz a la oscuridad** *(el segundo conjunto de curvas, desde arriba)*. **El total de la modulación que aprecia el rabdoma** *(las terceras estructuras cilíndricas, desde arriba)* **está en función del campo de visión del rabdoma $\Delta\rho$ y del espaciado de las bandas$\Delta\theta$. A la izquierda, un rabdoma estrecho ve un dibujo en movimiento de bandas anchas; el campo de visión del rabdoma es una cuarta parte del período de las bandas ($\Delta\rho={}^1/_4\Delta\theta$). La sensibilidad del rabdoma a lo largo del campo de visión queda representada en forma de picos de las curvas debajo de las estructuras cilíndricas del rabdoma. Conforme las bandas anchas van atravesando el estrecho campo de visión, el rabdoma va apreciando una modulación de gran amplitud** *M (curva de abajo, a la izquierda)*. **En el centro, un rabdoma estrecho percibe bandas estrechas en movimiento ($\Delta\rho = {}^1/_2\,\Delta\theta$) y aprecia una modulación rapida de amplitud moderada. A la derecha, un rabdoma ancho detecta bandas estrechas en movimiento ($\Delta\rho = \Delta\theta$) y distingue un perfil gris parpadeánte, que podría estar por debajo del nivel del ruido óptico para cualquier condición, excepto para la iluminación más intensa o para los ojos dotados de facetas mayores. Las líneas discontinuas simbolizan el movimiento en el interior del ojo.**



LA AMPLITUD DE LA MODULACION RELATIVA *M*, apreciada por el rabdoma como un dibujo listado que se mueve a través de su campo de visión, aparece aquí como una función de la razón entre la anchura del campo de visión y el espaciado de las bandas. Cuando el campo es igual al espaciado de las bandas ($\Delta\rho = \Delta\theta$), la modulación es de sólo un pequeño tanto por ciento, que puede bastar para la visión con luz solar intensa. Cuando el campo de visión es igual a la mitad del espaciado ($\Delta\rho = {}^1/_2\Delta\theta$), la modulación es de un 40 por ciento, lo que se necesitaría en niveles de luz inferiores para vencer el ruido de los fotones. De aquí que los receptores que han de trabajar a niveles de luz bajos tengan un campo muy extenso ($\Delta\rho$), y deban poseer un diámetro grande (d).

ceptor es igual a la diferencia de la intensidad máxima y la mínima, dividida por la suma de las intensidades máxima y mínima. Así una modulación especial *m*, en el contraste entre luz y oscuridad, en el campo visual se convierte por el movimiento y el receptor en una modulación temporal *M;* su amplitud relativa depende de la anchura del campo de visión, del espaciado de las rayas y (a velocidades significativas) de la velocidad angular del movimiento relativo *(véase la ilustración de la página 89)*.

Un receptor con un amplio campo de visión $\Delta\rho$ captará más luz y tendrá un sistema para conocer la razón con el ruido (relación señal-ruido) mejor que otro receptor con un campo de visión estrecho (por un factor de $\Delta\rho^2$); ahora bien, un receptor con un campo de visión grande será también menos sensible a la modulación producida por rayas finas. De la misma manera, un omatidio con una faceta de gran diámetro *D* experimentará una modulación más firme por un factor D^2. Es esencial una transmisión eficaz de la modulación al interior del ojo, en particular cuando la intensidad de la luz ambiente es baja. Por consiguiente, si puede determinarse el nivel de la señal luminosa necesario para activar los receptores por encima del nivel de ruido óptico, se puede llegar a la relación óptima entre el diámetro de las facetas y el ángulo interomatidial para insectos que se muestren activos en diferentes niveles de iluminación.

Estamos ante una fuente extensa de luz, de iluminación constante, cuando emite por término medio un número de $\bar{N}$ fotones por unidad de tiempo. Los fotones llegan al rabdoma más o menos al azar; en cualquier instante el rabdoma puede recibir un número de fotones superior al promedio y en el instante siguiente puede recibir un número inferior a este promedio. La diferencia entre el número de fotones que se reciben en un momento dado con el número medio es el ruido inevitable; se define por la desviación típica con respecto al número de fotones promedio y, para la luz, es igual a la raíz cuadrada del promedio de fotones, o $\sqrt{\bar{N}}$.

La señal absoluta que el rabdoma recibe es la amplitud relativa de modulación *M* multiplicada por el número promedio de fotones captados. M $\bar{N}$. La razón de la señal al ruido es, por tanto, M $\bar{N}/\sqrt{\bar{N}}$, que equivale a $M\sqrt{\bar{N}}$. Esta razón de la señal al ruido debe ser suficientemente grande para que la señal emerja por encima del ruido, y el omatidio no responda a falsas alarmas. Si $\bar{N}$ decrece hasta el límite, *M* debe aumentar correlativamente. Si el rabdoma está recibiendo de una fuente luminosa puntual un promedio de, digamos, cuatro fotones por unidad de tiempo, el ruido de este número de fotones es dos fotones en más o en menos; así pues, la modulación *M* en el número promedio de fotones (la señal) debe ser superior al 50 por ciento del número promedio de fotones para que esta modulación se sitúe por encima del ruido. Con luz intensa, sin embargo, la modulación puede ser inferior mientras el producto $M\sqrt{\bar{N}}$ sea de alrededor de 1. Si el rabdoma está recibiendo un promedio de, digamos, 100 fotones por unidad de tiempo, el ruido será de 10 fotones en más o en menos, de suerte que la modulación *M* en el número promedio de fotones sólo necesita ser del 10 por ciento, aproximadamente.

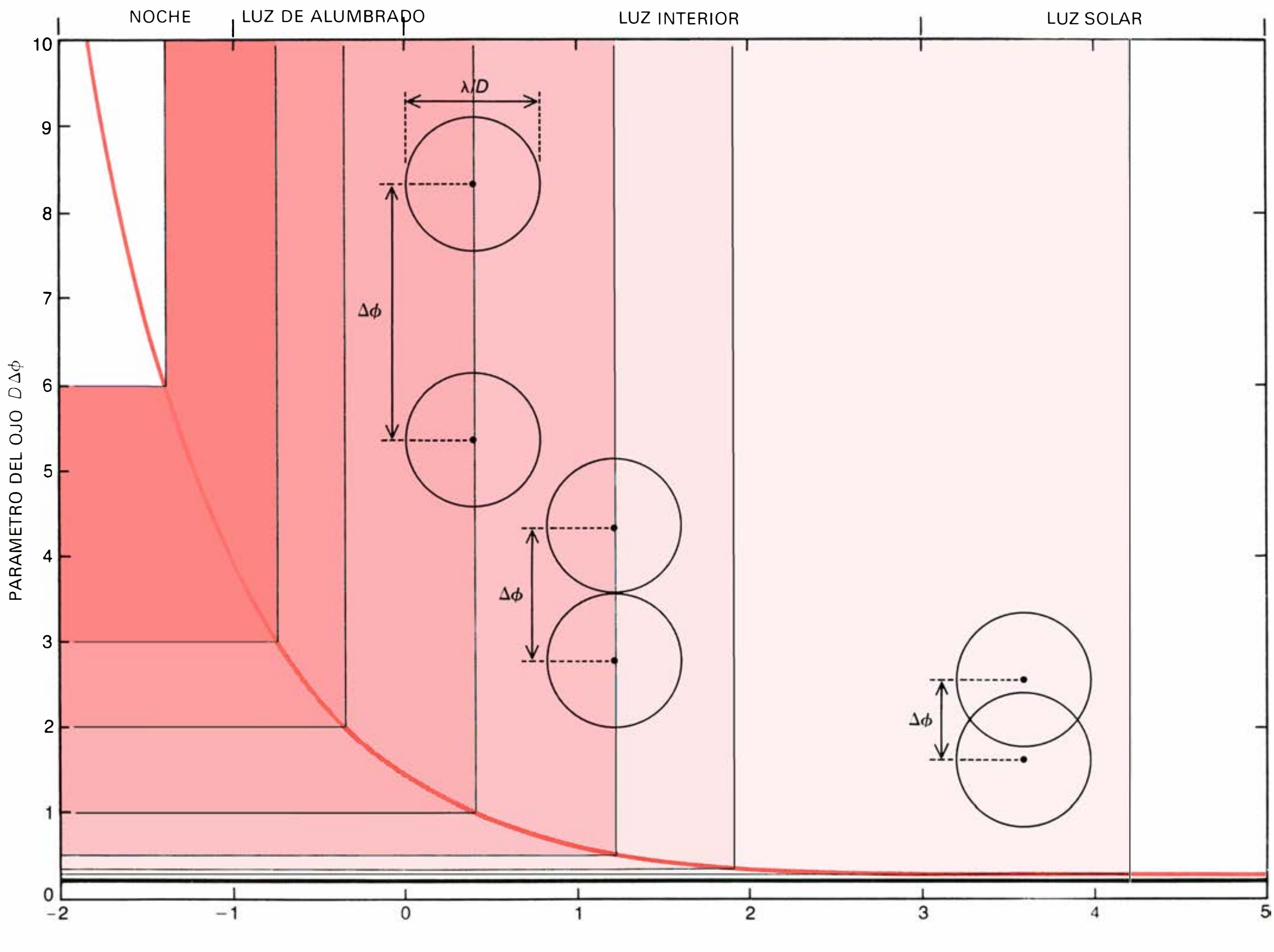
Hay ahora suficientes ecuaciones para calcular el parámetro del ojo $D\,\Delta\,\Phi$ para cualquier valor de la intensidad de luz del ambiente; los resultados se observan en forma de curva suave en la ilustración de la página 91 (tomado del reciente trabajo de Allan Snyder, Doekele Stavenga y Simon Laughlin). Con esta curva se puede calcular para cualquier intensidad de luz ambiental la cantidad con la que los círculos de los mapas del ojo se corresponderán con la mejor visión. Para un ojo limitado por la difracción, que tenga un parámetro del ojo de 0,3 o menos, los círculos de diámetro λ/D de los mapas del ojo se superpondrán uno sobre el otro casi en su mitad, y la amplitud real de campo $\Delta\rho$ podrá ser dos veces el ángulo interomatidial $\Delta\Phi$. Para los ojos de insectos que tengan un parámetro ocular de 0,5 la amplitud del campo sera de alrededor de 1,3 veces mayor que el ángulo interomatidial; para los que tengan un parámetro ocular de 2, la amplitud del campo $\Delta\rho$ será sólo las tres cuartas partes del ángulo interomatidial $\Delta\,\Phi$. Si estas dimensiones concuerdan o no realmente con las existentes en los ojos de insectos, es algo que sólo nos lo podrá confirmar un detallado estudio de gran variedad de diferentes tipos de insectos.

La sensibilidad del ojo del insecto puede aumentarse en un factor D^2 al aumentar la apertura *D* del cristalino del omatidio o en un factor $(d/f)^2$ al disminuir la distancia *f* desde el punto nodal posterior al rabdoma. En uno y otro caso, se amplía el campo de visión. En los ojos compuestos de las libélulas grandes, saltamontes, mántidos, avispas y coleópteros (y también en los ojos de muchos

crustáceos) la curvatura de la lente es hacia el interior, de modo que el punto nodal posterior queda un poco más unido al rabdoma y el ojo puede funcionar igualmente bien si se halla dentro del agua o fuera de ella. La superficie del ojo es entonces casi lisa y vidriosa. En muchos ojos de insectos se encuentra un pigmento absorbente con luz fuerte alrededor del rabdoma. Este pigmento absorbe luz en los ojos que tienen que funcionar a lo largo de una amplia gama de intensidades. A oscuras, el pigmento se retrae hacia atrás, aumentando la sensibilidad del rabdoma y ampliando a veces el campo de visión del omatidio. En algunos insectos, la adaptación a bajos niveles de luz se complementa con el movimiento del receptor hacia la lente. Los ojos compuestos que se adaptan a la oscuridad por este sistema son los que se conocen como ojos de cono móvil (o acono); se encuentran en las chinches de campo, mosquitos y algunos escarabajos.

Muchos insectos nocturnos y algunas polillas diurnas y escarabajos poseen una superposición de ojos con un sistema óptico completamente diferente, con el punto nodal posterior en el centro de curvatura de todo el ojo. La sensibilidad de este tipo de ojos está aumentada por tener los receptores próximos al punto nodal posterior. Incluso en otros insectos, el valor del parámetro del ojo $D \Delta \Phi$ en un punto del ojo es completamente diferente del valor en otra posición, lo que sugiere que el ojo es sensible a una gama amplia de intensidades. Podemos ver ahora el valor de conocer el parámetro del ojo para cualquier parte del ojo compuesto antes de proceder el estudio del ojo en otros sentidos.

Por ejemplo, consideremos el ojo de la mosca común. Cuando el parámetro del ojo $D \Delta \Phi$, medido a partir de mapas oculares, se compara con la intensidad de luz ambiental a la que la mosca es activa normalmente, vemos de inmediato lo que podría ser un desajuste: el ojo de la mosca está proyectado, según parece, para operar con una intensidad de luz unas 10.000 veces más baja que la que se encuentra a pleno sol. Este desajuste se



EL PARAMETRO OCULAR $D\Delta\Phi$ *(curva en color)* **tiene un valor óptimo diferente para distintos niveles de intensidad de luz del ambiente. Si la estructura de los ojos de los insectos estuviera determinada sólo por los límites de difracción de la luz, los ojos de estos artrópodos tendrían un parámetro ocular común de unos 0,25** *(línea horizontal negra que cruza la ilustración, abajo).* **Los círculos de** α/D **en el mapa del ojo deberían entonces superponer sus mitades (no indicado aquí). En cambio el parámetro ocular solamente se acerca al límite de difracción en las fóveas de insectos activos en presencia de luz solar extraordinariamente intensa** *(a la derecha).* **Además, el ojo del insecto suele poseer valores distintos del parámetro ocular en diferentes zonas, con la finalidad, al parecer, de que cada área sea lo más eficaz posible en la intensidad luminosa a la que debe trabajar. La candela por metro cuadrado** *(eje horizontal)* **es una unidad de intensidad luminosa.**

resuelve con sólo pensar que las moscas no sólo prefieren los lugares umbrosos sino que también utilizan los ojos cuando revolotean a gran velocidad. Las maniobras a gran velocidad provocan que el ojo pase por el medio ambiente a elevadas velocidades angulares, reduciendo el número eficaz de fotones de que dispondrá el fotorreceptor. Los giros a gran velocidad disminuyen la intensidad a que se percibe la luz en un factor de 1000, aproximadamente. Estas consideraciones no son ajenas a las razones del revuelo de las moscas, del encaramarse de las libélulas y de la permanencia inmóvil de la mantis religiosa.

Con la óptica del ojo compuesto *in mente*, podemos analizar algunos mapas de los ojos de los insectos preparados a partir de las mediciones de la pseudopupila. Cada círculo de los mapas oculares tiene un diámetro de $5\lambda/D$ radianes, tomando la longitud de onda λ como 0,5 micrómetros (luz verde). Cada círculo representa la anchura del círculo borroso de uno de cada cinco omatidios. El tamaño de los círculos es, pues, la llave para la resolución local de cada omatidio: cuanto más pequeño sea el círculo, mejor será la resolución. La distancia del centro de un círculo al centro del siguiente es cinco veces el ángulo interomatidial ($5\Delta\,\Phi$). La proximidad de los círculos es un índice de la densidad con la que el ojo del insecto capta el mundo visual y, por consiguiente, es un índice de la resolución visual en aquella región del ojo. La imbricación de los círculos muestra la importancia del parámetro del ojo $D\Delta\,\Phi$ en aquel lugar. Por ejemplo, donde los círculos se tocan, su diámetro λ/D es igual a su separación $\Delta\,\Phi$ y el parámetro del ojo vale de 0,5 micrómetros.

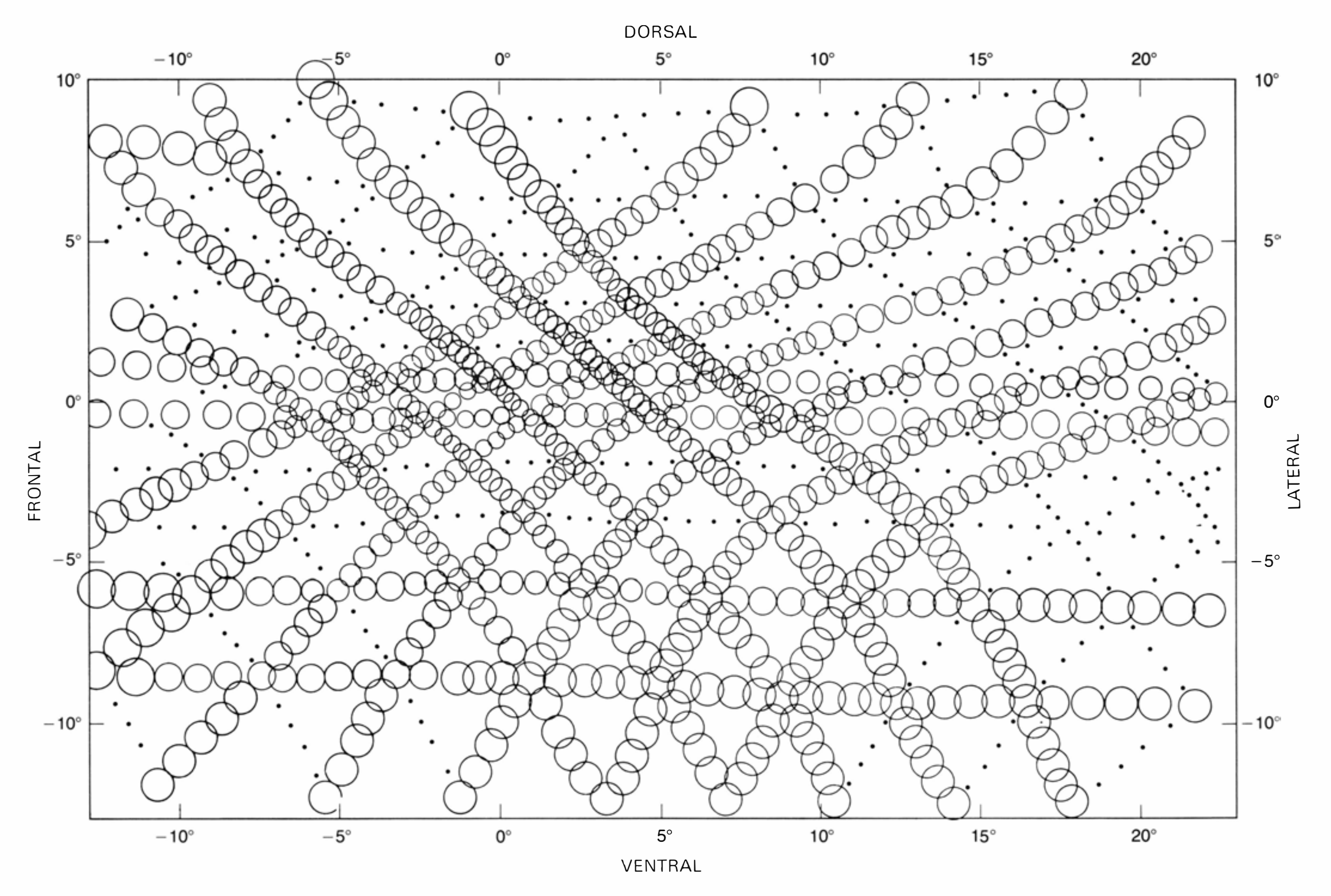
He compilado mapas de ojos de tipos que fueran representativos del grupo de los insectos y he encontrado una serie de formas generales.

Primero, no parece que haya insectos cuyos ojos estén destinados en principio a operar cerca del límite de difracción. Los primeros trabajos que sugirieron que los ojos de los insectos quedaban limitados por la difracción eran erróneos. También los ojos de una avispa *Bembix* de Australia, que se muestra activa en plena luz solar del desierto, tienen un parámetro del ojo de 32 sólo en el centro de la fóvea. Los ojos de los mántidos, que capturan presas que permanecen totalmente paradas, con luz intensa poseen un parámetro del ojo sorprendentemente grande, entre 0,5 y 2.

Segundo, la mayoría de los insectos que vuelan en intensa luz solar, como las abejas, mariposas y libélulas, poseen un parámetro del ojo de unos 0,5. Esta generalización habría de aplicarse con tiento, debido a que cualquier variación mínima del valor del parámetro del ojo cerca de esa cifra representa una gran diferencia en la intensidad óptima en la que el ojo puede funcionar de la manera mejor.

Tercero, los ojos de los insectos carnívoros parecen tener un parámetro del ojo mayor que los ojos de los insectos herbívoros del mismo hábitat, si bien los insectos carnívoros poseen también fóveas.

Cuarto, es frecuente encontrar una pendiente suave en el valor del paráme-



LA AVISPA *Bembix palmata* revela que los ejes ópticos de los omatidios del insecto *(puntos y centros de los círculos)* no siguen el mismo modelo hexagonal que las facetas del ojo. A lo largo del eje horizontal del mapa, forman cuadrados oblicuos que se hacen más oblicuos hacia el ecuador del ojo. Los círculos tienen un diámetro de λ/D. La gran superposición entre ellos indica que el parámetro ocular es de alrededor de 0,3. *Bembix* es uno de los pocos insectos para los que el parámetro ocular (el parámetro en la fóvea cerca de 0 grados, 0 grados en coordenadas angulares) se acerca muchísimo al límite de difracción. *Bembix* es activo en la luz solar cegadora del desierto australiano, de forma que el ruido de los fotones es bajo con respecto a la intensidad de la luz y se puede tender al límite de difracción. (Ilustración realizada por Type Systems Inc.)

tro del ojo, en dirección horizontal a través del ojo de un insecto, que va desde los valores más bajos cerca de la parte frontal del ojo hasta los valores más altos en la vecindad de los lados. Los gradientes del parámetro del ojo deben estar relacionados, presumiblemente, con el hecho de que, cuando el insecto está volando, los objetos situados lateralmente al ojo tienen una velocidad angular mayor y producen una modulación menor que los objetos situados delante del ojo. Las pendientes se distinguen perfectamente siguiendo una fila horizontal de facetas en los ojos de los saltamontes, himenópteros y libélulas, aunque también son característicos de mántidos, hormigas y de otros insectos que no poseen filas horizontales de facetas o que no tienen una rápida velocidad de avance.

Quinto, a lo largo de la región ecuatorial del ojo, la pseudopupila queda alargada verticalmente, ya que aquí el ángulo interomatidial entre las facetas de filas verticales es menor que el ángulo interomatidial entre las facetas de filas horizontales. Aunque el modelo hexagonal aparente de las facetas del ojo está poco modificado donde la pseudopupila es alargada, los ejes ópticos de las facetas quedan comprimidos verticalmente para formar un enrejado de rombos, de suerte que los ojos contemplan el entorno desde una formación de cuadrados oblicuos. Tal como hemos visto, estos modelos cuadrados capacitan al ojo para recoger el mundo visual con mayor eficacia que se logra con el modelo hexagonal. De todo lo cual se deduce la consecuencia interesante de que los ojos pueden usar las filas verticales de omatidios mejor que las filas horizontales en misiones que requieran un alto poder resolutivo.

Sexto, donde hay mucha superposición binocular, el ojo se dilata en un ángulo total más ancho, y el parámetro del ojo resulta entonces mayor que el esperado. Este es el caso de los mántidos.

Los mapas del ojo aportan asimismo abundante información sobre las fóveas de los insectos.

Primero, la fóvea se forma allí donde las filas de los ejes ópticos se juntan, de modo que la densidad de los omatidios que miran es mayor. Una fóvea no se forma nunca por un suplemento de filas de omatidios extras. Todo lo contrario: en una zona foveal tienen que ser suprimidas filas de facetas para dejar espacio a otras facetas mayores.

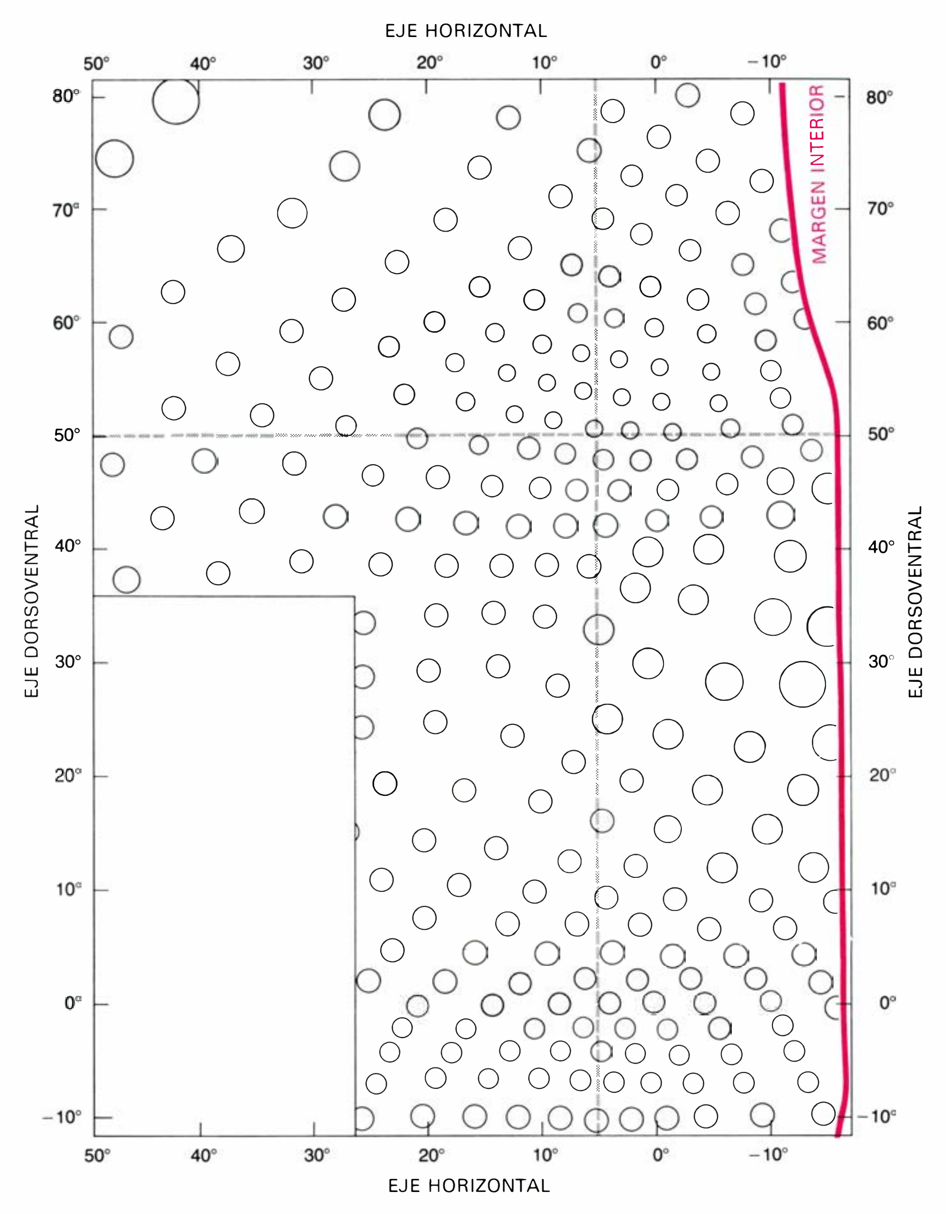
Segundo, los ejes ópticos se unen por inflexión de las filas de ejes: el que las facetas de una fila estén en línea recta no indica necesariamente que sus ejes ópticos se hallen en un mismo plano.

Tercero, en los mapas de ojos de algunos insectos depredadores, los círculos de diámetro λ/D se superponen en la fóvea más que en las áreas colindantes del ojo. Ello significa que, en la fóvea, el parámetro del ojo se hace menor: a su vez, esto implica que la fóvea se muestre como una región adaptada para ver con luz relativamente intensa. En ciertos insectos (*Austrogomphus*), el parámetro del ojo permanece constante a través de la fóvea.

Cuarto, las filas de facetas son siempre más regulares en la fóvea que en cualquier otra parte del ojo, lo que demuestra que la regularidad desempeña un papel importante en estas regiones del ojo en donde se necesita ver lo mejor posible.

Quinto, aunque las facetas de la fóvea normalmente son mayores que las de las otras zonas del ojo, no sucede así en el caso del saltamontes *Locusta* y del mántido *Orthodera*. Ambos insectos poseen una fóvea dirigida hacia adelante, donde el ángulo interomatidial es menor y donde las facetas tienen el mismo tamaño que las del resto del ojo. Una fóvea de este tipo mejora la acuidad visual, si bien posee campos de visión más estrechos y por ello menos sensibles.

Sexto, en la avispa de arena *Bembix*, que vuela con la luz extraordinariamente intensa del desierto australiano, el parámetro del ojo es muy bajo, de unos 0,31



EL MAPA DEL OJO DE LA LIBELULA NOCTURNA *Zyxomma* muestra el eje óptico de uno de cada cinco omatidios. *Zyxomma*, un insecto del sureste de Asia, sólo sale a comer en el crepúsculo. El mapa muestra la extraordinaria separación de los círculos (de diámetro $5\lambda/D$ radianes), con la fóvea superficial del ojo del insecto mirando oblicuamente hacia adelante y hacia arriba cerca de los 0 grados, 50 grados sobre los ejes, y cruzando los ejes por el eje nulo de abscisas *(abajo)*. Las líneas discretas muestran líneas en las que no hay errores causados por la traducción del mapa del ojo de la libélula nocturna desde la superficie curva del ojo a la superficie plana del papel.

en la fóvea, cerca del límite de difracción. Muchos insectos, sin embargo, poseen un parámetro del ojo notablemente grande también en la fóvea. Por ejemplo, el parámetro del ojo alcanza 1 en la fóvea de los *damselflies*, que buscan sus presas entre las cañas y debajo de los árboles. En cualquier caso, la fóvea es la zona de mayor poder resolutivo visual, por más que pueda ser una zona de baja sensibilidad o hallarse configurada para funcionar a la misma intensidad luminosa que el resto del ojo, teniendo facetas mayores.

Aunque puede ser cierto que las costumbres de los insectos están escritas en sus globos oculares, es difícil leer los signos correctamente, porque sabemos muy poco sobre cómo utilizan los ojos. Los ojos que miran hacia los lados, al igual que los de ciertos pájaros o el conejo, son, presumiblemente, sistemas de alerta. Las langostas y las moscas, que son muy sensibles a los movimientos a los lados de la cabeza y relativamente poco sensibles al movimiento en el frente, ofrecen un indicio de ello. La gran fóvea lateral que mira directamente a los lados en muchas libélulas grandes viene a complicar las cosas, tal vez tenga por misión ayudar a revolotear al animal por la superficie del agua. Los ángulos interomatidiales entre las facetas de esta fóvea tienen menos de 0,5 grados, y son los más pequeños que he encontrado en los insectos.

Los insectos que cazan a la vista precisan seguramente de una fóvea que mire hacia adelante. Existen dudas, sin embargo, sobre si la mayoría de fóveas de los insectos, como las de los vertebrados, trabajan mejor con luz más fuerte. Si sucede así, las libélulas que cazan avispas y mántidos pueden sufrir una disminución de la acuidad visual al anochecer, como le ocurre al hombre.

Los ojos de los mántidos son desconcertantes en muchos aspectos. En la mayoría de ellos, la fóvea se destaca manifiestamente. Ya en 1971 J. C. Barros-Pita y H. Maldonado, en Venezuela, vieron que el área inmediatamente circundante de la fóvea, como se ha dicho aquí (no el centro de la fóvea), es el área esencial que permite al insecto ver lo bastante bien como para capturar la presa. Por otra parte, excepción hecha de una larga lista de mántidos para los que he construido mapas del ojo, ninguno posee círculos de λ/D que superponga exteriormente a la región de la fóvea. Y lo que es más notable, los ojos de los mántidos mayores, que tienen facetas de unos 50 micrómetros que atraviesan el centro de la fóvea, tienen un parámetro del ojo superior a 2; por consiguiente, la separación de centro a centro de los círculos de los mapas oculares es mayor a dos veces su diámetro λ/D. La única explicación plausible es que el ojo del mántido tiene 45 grados de superposición binocular en cada ojo, de suerte que los ojos se dilatan en un campo de visión anormalmente ancho.

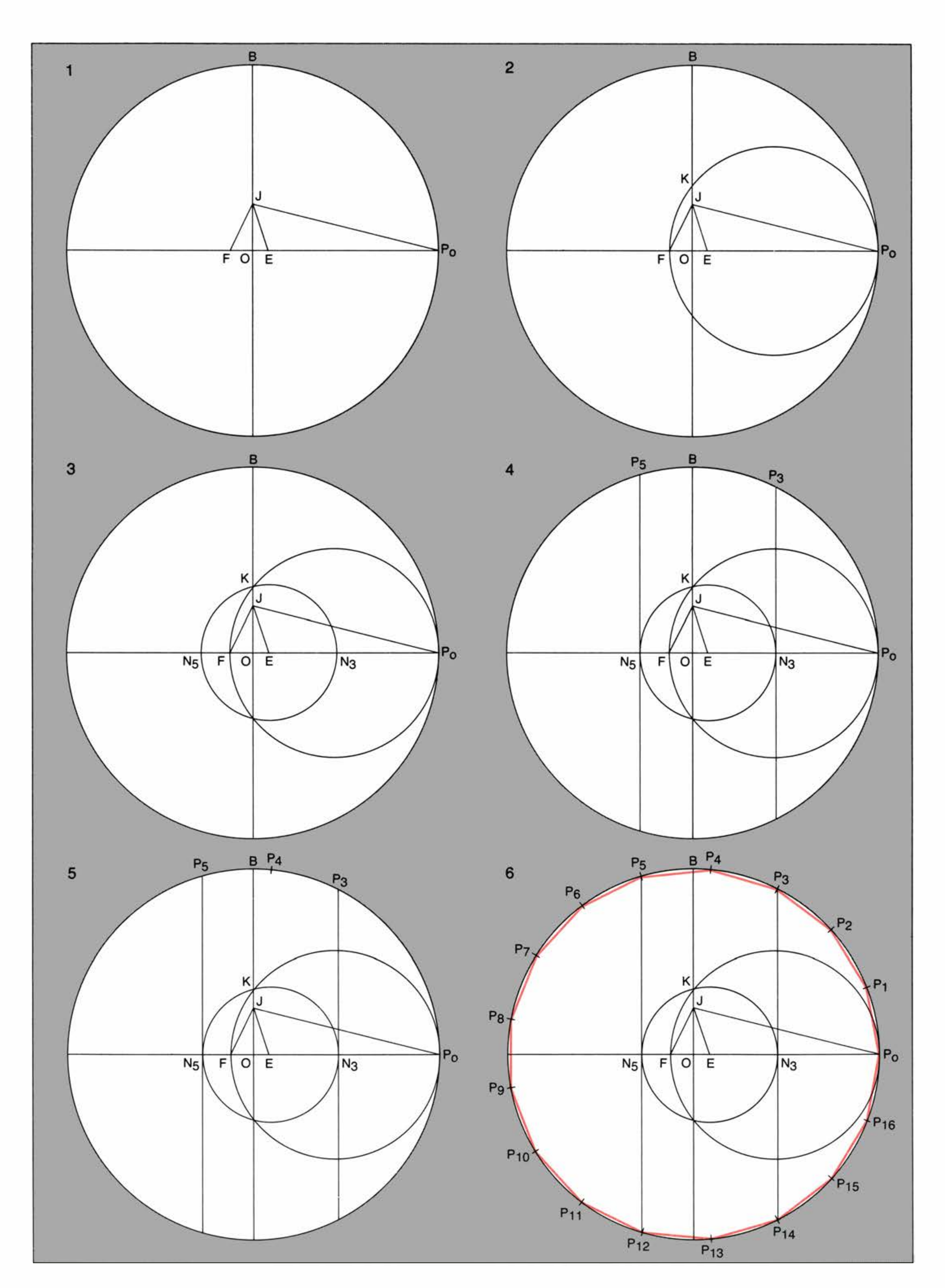
En principio, el hecho de que las libélulas tengan fóveas que miran oblicuamente hacia adelante y hacia arriba resulta también desconcertante. Puede ser interesante recordar aquí que un helicóptero que vuela a toda velocidad se inclina hacia adelante, de modo que la rotación de las palas de su hélice, que lo levanta, es responsable también de su propulsión hacia adelante. Las libélulas guardan cierto parecido con los helicópteros; cabe el que sus fóveas, dirigidas hacia arriba, estén apuntando directamente a la presa que persiguen por el aire. Lo que necesita un insecto para prever la posición de algo que está cayendo o para acercarse a las presas por debajo es exactamente una fóvea que mire hacia arriba. La parte superior del ojo de la libélula es, además, especialmente sensible a las longitudes de onda corta, lo que sugiere que el ojo aprecia los contrastes antípodas y tiene su mejor poder resolutivo al mirar algo en el azul del cielo.

La relación teórica entre el valor del parámetro del ojo y la verdadera intensidad de luz a la que cada parte del ojo compuesto puede funcionar sirve para reunir una amplia cantidad de datos. La teoría está basada en tres supuestos fundamentales: primero, que el ojo compuesto es una disposición de receptores que procesa la información simultáneamente a través de canales paralelos; segundo, que las facetas y rabdomas del ojo deben ser lo suficientemente anchos como para captar el número de fotones que permita percibir la modulación que está por encima del ruido óptico; tercero, que un modelo listado resuelto por cualquier receptor individual puede ser visto también como un modelo por todo el ojo como una unidad.

Tras una incursión en los estudios sobre la visión de los insectos, se podría objetar que no hay pruebas de que el último supuesto sea cierto, o que un modelo fino sea recompuesto realmente por cualquier ojo compuesto. La prueba existe. Primero, numerosos ensayos con abejas que habían sido amaestradas para responder ante ciertos modelos muestran que hay dos aspectos en todo modelo que son importantes para una abeja: la cantidad de ondulaciones que el modelo o dibujo forma al pasar a través del ojo compuesto, y los ángulos y colores del perfil contrastante del mismo. Los dibujos que el ojo humano ve como muy diferentes son fácilmente confundidos por la abeja amaestrada si los dibujos producen una cantidad de ondulaciones similar. Segundo, para que un insecto perciba la velocidad y dirección de un dibujo que se mueve, lo importante no es el dibujo en sí mismo sino la secuencia aparente en sucesivos ejes a través del ojo.

Es difícil proyectar un experimento que pruebe definitivamente si el insecto reconstruye o no en su cerebro el dibujo y, por tanto, si lo interpreta como un dibujo real. Sin embargo, resulta fácil demostrar que un insecto responde a un movimiento muy pequeño de un objeto grande y que responderá, especialmente bien, a un pequeño movimiento de un objeto pequeño. En muchos insectos pequeños, que no tienen espacio en la cabeza para alojar grandes ojos, el ángulo interomatidial puede ser de 10 grados; muchos y esmerados estudios han demostrado que estos insectos pueden responder con detalle con la décima parte del ángulo interomatidial. Parte del secreto radica posiblemente en que algunos insectos pueden hacer uso de la vibración del ojo para escudriñar el mundo visual y obtener así una imagen más acabada como un modelo en el tiempo en lugar de serlo en el espacio. Eso iría contra el supuesto de que la recogida de imágenes debe ser simultánea.

En parte, las objeciones que se pueden levantar contra los supuestos relativos a la teoría óptica que he expuesto aquí se invalidan las unas con las otras. Por ejemplo, ver el movimiento precisa que el insecto sea sensible a la computación temporal justa de las pequeñas modulaciones en cada eje visual sobre una región del ojo. Ver la dirección que toma un perfil contrastante requiere que el insecto sea sensible a la fase exacta de las componentes sinusoidales (listadas) del dibujo. Ver el movimiento relativo de un objeto pequeño y oscuro en cualquier dirección sobre un fondo moteado que también se mueve, como un insecto en vuelo cazando, requiere toda la información que el ojo pueda obtener. Sospecho que si examinamos el comportamiento visual de los insectos más de cerca encontraremos que hay un premio para el que recoja toda la información posible. La forma de los ojos compuestos más avanzados sugiere que están proyectados para recoger la máxima cantidad de información que puedan en batalla tenaz contra el ruido de los fotones.


1
B
J
F
O
E
Po
2
B
K
J
F
O
E
Po
3
B
K
J
N5
F
O
E
N3
Po
4
P5
B
P3
K
J
N5
F
O
E
N3
Po
5
P5
B
P4
P3
K
J
N5
F
O
E
N3
Po
6
P5
B
P4
P3
P6
P2
P7
P1
K
J
P8
N5
F
O
E
N3
Po
P9
P16
P10
P15
P11
P14
P12
P13

Gauss

Niño prodigio, llegó a ser el principal matemático de su época. Se desenvolvió con igual soltura en las abstracciones de la teoría de números y los complejos cálculos astronómicos como en los aspectos más prácticos de la física aplicada

Ian Stewat

La Matemática es la reina de las ciencias, dijo en cierta ocasión Carl Friedrich Gauss; su propia vida sirvió de ejemplo a este aforismo. Considerado por todos, al par de Arquímedes y Newton, como uno de los matemáticos más capaces de todos los tiempos, Gauss se interesó tanto por la teoría como por las aplicaciones, y sus contribuciones van desde la más pura teoría de números hasta los problemas prácticos de astronomía, magnetismo y topografía. Realizó profundos descubrimientos en todas las ramas de la matemática en las que trabajó, introdujo ideas y métodos nuevos y estableció los cimientos de investigaciones posteriores. Da medida de su talento el que este año, bicentenario de su nacimiento, todavía sean fecundas muchas de sus ideas.

Gauss fue, en muchos aspectos, una personalidad contradictoria y enigmática. Hijo único de padres de clase obrera, ascendió hasta la cumbre de la matemática de su época; vivía, sin embargo, con modestia, y rehuía ser conocido públicamente. De suaves modales, era un hombre distante, políticamente reaccionario y frecuentemente testarudo, que tan sólo pedía poder continuar, sin perturbaciones, su trabajo de creador. Siempre dispuesto a reconocer el talento matemático allí donde estuviese, por encima de prejuicios de sus contemporáneos, dejó en el olvido a varios de los mejores jóvenes matemáticos de su tiempo, en especial, a János Bolyai, uno de los pioneros de la geometría no-euclídea, lo que tuvo consecuencias poco afortunadas.

Un aspecto especialmente llamativo del carácter de Gauss fue su rotunda negativa a presentar parte alguna de su trabajo que no creyera haber pulido hasta la perfección. Ningún resultado, por importante que fuese, se publicó hasta que él no lo consideró completo y terminado. Tan elaboradas son sus demostraciones matemáticas, que el camino que siguió para obtenerlas se esfuma completamente. Sus trabajos publicados tienen la calidad, austera e inabordable, la gracia y la elegancia clásicas. Muchas de sus ideas más fecundas no aparecen explícitamente en su obra impresa, y es preciso inferirlas reconstruyendo los pasos que debieron conducirle a su descubrimiento. De resultas, muchas nociones importantes no han visto la luz del día hasta ser independientemente descubiertas por otros.

Gauss publicó en vida unos 155 títulos, y dejó tras sí gran cantidad de trabajos inéditos. En esta breve exposición abordaré algunos de sus más importantes, y fecundos, descubrimientos, e intentaré, en pequeña medida, explicar cómo llegó a obtenerlos.

Gauss nació el 30 de abril de 1777 en Brunswick (perteneciente en la actualidad a Alemania Federal). Su padre pasó por diversos oficios: jardinero, guarda de canales y albañil. Su hijo lo describiría más tarde como "hombre totalmente honesto, y, en muchos aspectos, estimable y genuinamente respetable, pero en casa... dominante, grosero y rudo". La madre de Gauss, hija de un cantero, era mujer inteligente y de recio carácter. Su hermano Friedrich tuvo importante papel en la vida del joven Gauss. Trabajaba de tejedor de damascos finos, pero su campo de interés era insólitamente amplio. Pasó mucho tiempo animando a Gauss y aguzando su espíritu crítico.

Hay entre los grandes matemáticos tantos que hayan mostrado talento matemático en su infancia como quienes no mostraron ninguno hasta mucho más tarde. Gauss ha sido, incuestionablemente, el más precoz de todos ellos. Solía decir, bromeando, que había aprendido a contar antes de aprender a hablar; muchas anécdotas atestiguan sus extraordinarias dotes. Se cuenta que un día, antes de cumplir los tres años, su padre estaba preparando la paga semanal de los obreros a su cargo, sin darse cuenta de que su hijo observaba el proceso con gran interés. Al finalizar sus cálculos, Gauss padre quedó sorprendido al oír una vocecita que le decía: "Padre, la cuenta está mal hecha. Debería dar...". Al repasar los cálculos comprobó que la cifra dada por el niño era la correcta. Lo notable de esta historia es que nadie le había enseñado aritmética.

Otras anécdotas refieren la continua precocidad que Gauss demostró en la escuela. A los 10 años fue admitido en la clase de aritmética. El maestro propuso un problema del tipo siguiente: hallar la suma de $1+2+3+...+100$, donde hay 100 números, y la diferencia entre cada sumando y su siguiente es siempre igual a uno. Hay un método sencillo para realizar tales sumas, que el maestro conocía, pero los escolares no.

Se tenía la costumbre de que el primer muchacho que resolviese un problema dejase su pizarra con el resultado sobre

LA CONSTRUCCION GEOMETRICA de un polígono regular de 17 lados, usando sólo regla y compás, fue el primer descubrimiento de este tipo desde los tiempos de Euclides; marcó, en 1796, el comienzo de la carrera matemática de Gauss, cuando contaba 18 años. En la página opuesta se ofrece una versión simplificada de su construcción, preparada por H. W. Richmond en 1893. Se procede del modo siguiente: (1) Se traza una circunferencia de centro O y radio OP_0 de longitud arbitraria. Se traza la recta OB, perpendicular a OP_0. Se determina un punto J a la cuarta parte del recorrido OB. Se halla un punto E tal que el ángulo OJE sea cuarta parte del ángulo OJP_0 (lo que puede hacerse mediante doble bisección de este ángulo). Se determina un punto F tal que el ángulo FJE mida 45 grados. (Puede obtenerse por bisección de un ángulo recto). (2) Se construye una circunferencia de diámetro FP_0. Esta circunferencia corta a OB en el punto K. (3) Se traza otra circunferencia de centro E y radio EK. Esta circunferencia define los puntos N_5 y N_3. (4) Se trazan las rectas N_3P_3 y N_5P_5, perpendiculares a OP_0. (5) Se traza la bisectriz del arco P_3P_5, a fin de obtener el punto P_4. (6) Se lleva sucesivamente la cuerda P_4P_5 sobre la circunferencia, a partir de P_0. Los puntos obtenidos se unen por segmentos rectilíneos para formar el polígono.

la mesa del maestro, el siguiente en terminar dejase la suya sobre la primera, y así sucesivamente. Apenas el maestro había terminado de plantear el problema cuando Gauss puso su pizarra sobre la mesa. "Ahí lo tiene", dijo Gauss. Durante toda la hora siguiente permaneció cruzado de brazos, recibiendo ocasionalmente alguna escéptica mirada del maestro, mientras los demás alumnos bregaban con tan larga suma. Al final de la hora, el maestro examinó las pizarras. En la de Gauss había tan sólo un número. Ya en su vejez, a Gauss le encantaba contar cómo, entre todas las respuestas, la suya fue la única correcta.

Es preciso decir en honor del maestro que éste quedó tan impresionado que compró con su propio dinero un libro de aritmética y se lo regaló a Gauss, quien lo devoró rápidamente. Gauss tuvo también la suerte de que el ayudante de aritmética del maestro, un joven de 17 años llamado Johann Martin Bartels, fuera un apasionado de las matemáticas, con lo que ambos pasaron muchas horas estudiando juntos. Al llegar al teorema del binomio, que enuncia que, para todo número n, la expresión $(1+x)^n$ es una serie, y que cuando n no es entero positivo esta serie es infinita, Gauss quedó descontento de la falta de rigor del libro que el maestro le había dado, y construyó una demostración. A pesar de no ser más que un escolar, fue el primer matemático que prestó seria atención a los problemas creados por los infinitos. Pocos niños prodigio en matemáticas van más alla de una gran facilidad de cálculo, pero el talento de Gauss alcanzaba, claramente, los más elevados dominios del pensamiento humano.



RETRATOS de Gauss *(en primer término)* **y del físico Wilhelm Weber (1804-91), con quien colaboró en muchos experimentos prácticos sobre magnetismo y telegrafía. La inscripción griega de la orla dice: "Dios hace aritmética"; la inscripción latina del lado derecho significa: "El fin corona la obra". La cita en griego de la parte inferior, que está tomada de** *La República,* **del filósofo ateniense Platón, se traduce por: "Los que tienen las antorchas se las pasarán a otros".**

A la edad de 14 años, Gauss fue presentado al Duque de Brunswick, quien había oído hablar de su reputación y se convirtió en protector suyo. Al año siguiente, Gauss entró en el Collegium Carolinum, de Brunswick, donde estudió, y pronto dominó, las obras de Newton, Leonhard Euler y Joseph Louis Lagrange. A los 19 años, había descubierto por sí solo, y demostrado, un notable teorema de la teoría de números conocidos como ley de reciprocidad cuadrática (del que hablaré más tarde). Para poder apreciar cuán importante fue esto, hay que darse cuenta de que anteriormente Euler había ya descubierto el teorema, y tanto él como Adrien Marie Legendre fracasaron en sus intentos por demostrarlo.

Cuando Gauss abandonó el Collegium Carolinum, en octubre de 1795, para ir a estudiar a la Universidad de Göttingen, se vio en el dilema de decidirse entre las matemáticas y su otro gran amor: el estudio de las lenguas antiguas, donde no era menos brillante. Se decidió el 30 de marzo de 1796, tras realizar uno de los descubrimientos más sorprendentes de la historia de las matemáticas.

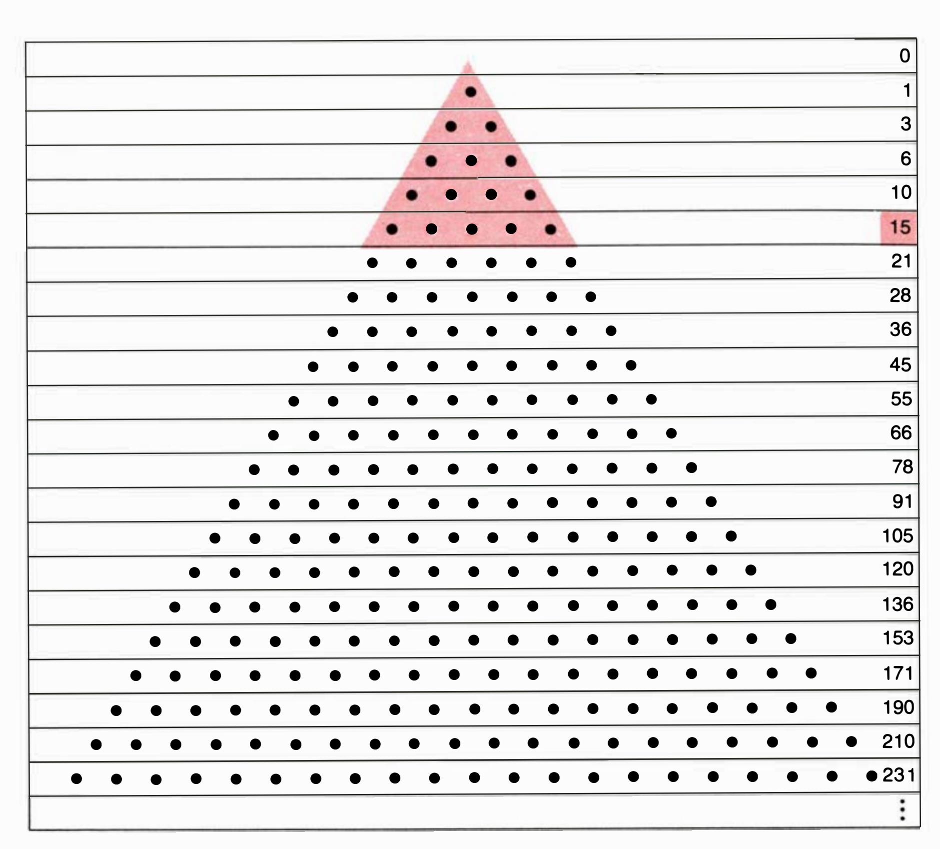
Para poder contemplar este resultado con alguna perspectiva, retrocedamos dos milenios, hasta la Grecia clásica. La principal contribución griega a las matemáticas se halla en la floreciente escuela geométrica ligada a los nombres de Pitágoras, Eudoxo, Euclides y Arquímedes. Probablemente fueron los griegos los primeros en reconocer la importancia del rigor de las demostraciones matemáticas; y al perseguir este rigor impusieron cierto número de restricciones a los métodos de demostración. Una de ellas, que en las construcciones geométricas tan sólo podrían utilizarse regla y compás. En efecto, las únicas líneas geométricas permitidas eran la recta y la circunferencia.

Euclides demostró que se pueden construir, con regla y compás, polígonos regulares de tres, cuatro, cinco y 15 lados, así como todos los deducidos de los anteriores por bisección reiterada de sus lados. Estos eran, sin embargo, todos los polígonos regulares que los griegos sabían construir; no conocían ningún método geométrico para construir polígonos de siete, nueve, 11, 13, 14 y 17 lados, por ejemplo. Durante los 2000 años siguientes parece que nadie llegó a sospechar que sería posible construir alguno de estos otros polígonos. El resultado de Gauss consistió en dar una construcción

del polígono regular de 17 lados, que inscribió en una circunferencia usando tan sólo regla y compás. Además, caracterizó exactamente los polígonos que podrán construirse por este método: el número de sus lados ha de ser potencia de 2 (2^n) o bien, potencia de 2 multiplicada por uno o más números primos impares distintos del tipo llamado "números primos de Fermat" (en honor de su descubridor, Pierre de Fermat). Un número primo es aquel que no puede dividirse exactamente por ningún número, excepto por sí mismo y por la unidad; un número primo de Fermat tiene la propiedad adicional de ser una unidad mayor que 2 elevado a una potencia de 2, o sea, $2^{2^n}+1$. Los únicos números primos de Fermat conocidos son 3, 5, 17, 257 y 65.537. Tenemos, así pues, el notable resultado de que, a pesar de ser posible construir con regla y compás polígonos regulares de 17 lados, no es posible hacerlo así para polígonos de siete, nueve, 11, 13 y 14 lados.

Gauss demostró este teorema (cuando contaba 18 años) combinando un razonamiento algebraico con otro geométrico. Demostró que construir un polígono de 17 lados es equivalente a resolver la ecuación $x^{16}+x^{15}+\ldots+x+1=0$. Como 17 es primo y 16 es potencia de 2, resulta que esta ecuación puede reducirse a una serie de ecuaciones de segundo grado (ecuaciones de la forma $ax^2+bx+c=0$, siendo a, b y c números dados, y x el valor a determinar). Como se había ya demostrado que las ecuaciones de segundo grado pueden resolverse mediante construcciones con regla y compás, la demostración está completa. Aparte de la importancia que esta demostración tuvo para inducir a Gauss a seguir la carrera de matemáticas, la demostración es notable porque constituye el primer ejemplo real (además de la introducción de coordenadas, debida a Descartes) de una técnica que, a partir de entonces, se ha convertido en una de las más fecundas de las matemáticas: trasladar un problema desde su dominio inicial (la geometría, en este ejemplo) a otro (álgebra) y resolverlo en este último.

Gauss escribió en cierta ocasión que, a la edad de 20 años, estaba tan sobrecargado de ideas matemáticas que no tenía tiempo de consignar sino una pequeña fracción de ellas. Muchas de las que pudo desarrollar aparecieron después en sus *Disquisitiones Arithmeticae*, publicadas en 1801 cuando contaba 24 años. Puede decirse que esta obra hizo por la teoría de números lo que Euclides por la geometría: organizó conocimientos dispersos sobre el sistema de los números



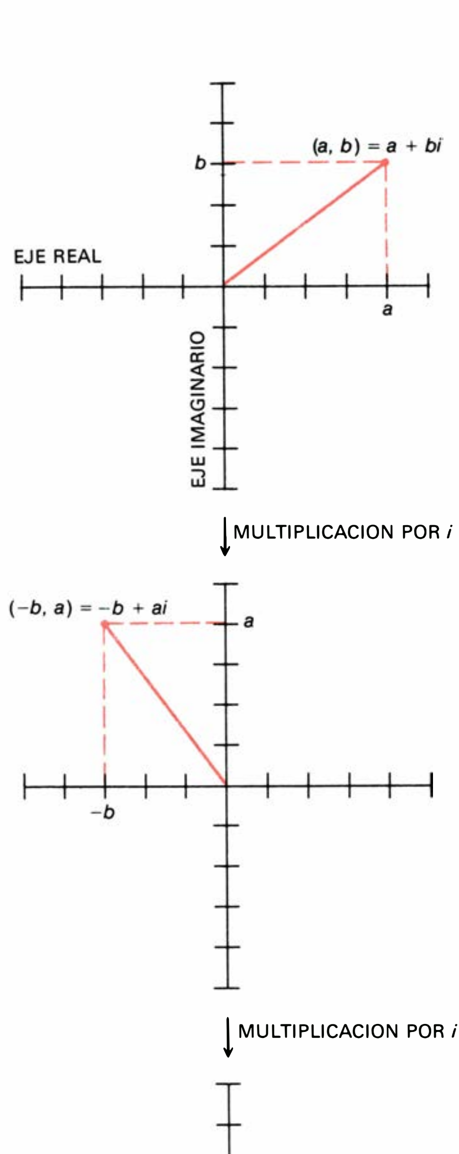
LOS NUMEROS TRIANGULARES son números de la forma $n(n+1)/2$, siendo n un entero positivo cualquiera. Puede también representarse mediante una disposición triangular de puntos. En sus *Disquisitiones Arithmeticae*, publicada en el año 1801, cuando tenía 24 años, Gauss demostró que todo entero positivo es suma de tres números triangulares. El 10 de julio de 1796 anotó el descubrimiento en su diario, con una críptica inscripción: ¡Eureka! num $=\Delta+\Delta+\Delta$.

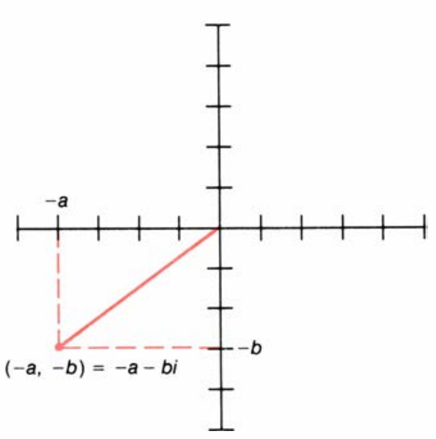
enteros, que Gauss complementó con algunas de sus más profundas ideas propias. Gauss fundamentó su teoría sobre la noción de números congruentes, que se definen como dos números a y b que tienen el mismo resto al dividirlos por un número m dado. Dos números cualesquiera que satisfagan esta condición se llaman "congruentes módulo m", siendo m un número fijo dado, llamado módulo. Por ejemplo, al dividir por 7 los números 16 y 23 ambos tienen el mismo resto, y así pues, son congruentes módulo 7. Siete y 9 son congruentes módulo 2, pues ambos tienen resto 1 al dividirlos por 2. Evidentemente, dos números pares y dos números impares cualesquiera serán siempre congruentes módulo 2.

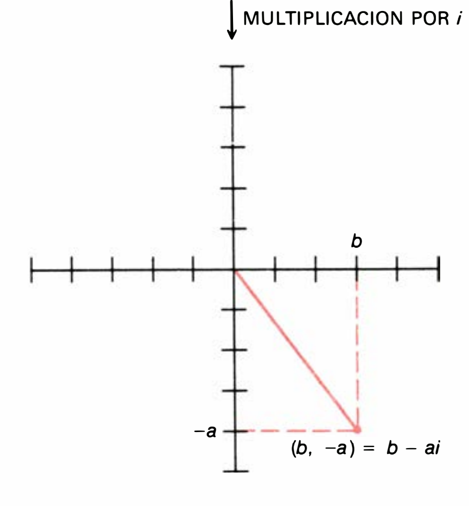
Gauss señaló también la posibilidad de realizar una aritmética de números congruentes. Demostró que, para cualesquiera enteros a, b, c y d, siendo módulo un número m, a es congruente con b y c es congruente con d, entonces $a+c$ es congruente con $b+d$ y ac es congruente con bd. Así pues, podemos sustituir enteros por números congruentes en la aritmética sin incurrir en contradicciones. Nos encontramos, sin embargo, con algunas sorpresas, como, por ejemplo, que módulo 3, $1+1+1$ es congruente con 0.

Esta aritmética de números congruentes se enseña en muchos cursos de "matemática moderna", con el nombre de aritmética modular o de cálculos de congruencias. Me pregunto cuántos profesores saben dónde se originó esta teoría y con qué objeto fue desarrollada por Gauss, quien la necesitaba como instrumento de demostración de profundos y difíciles teoremas. Quizás el más valioso de todos ellos, y con certeza, el favorito de Gauss, sea la ley de reciprocidad cuadrática, que Gauss llamó teorema áureo. Se hace necesaria ahora alguna terminología adicional.

Gauss definió en primer lugar los "residuos cuadráticos", diciendo que si m es un entero positivo, y a es un entero que no tiene divisores en común con m, entonces a es residuo cuadrático de m si es congruente, módulo m, con un cuadrado perfecto. También podemos enunciarlo del modo siguiente: Si a es residuo cuadrático de m, se puede hallar al menos un x cuyo cuadrado dividido por m dé resto a. Así, 13 es residuo cuadrático de 17, porque el enunciado "x^2 es congruente con 13 módulo 17" es verdadero si x toma el valor 8 (entre otros posibles). Gauss demostró que si p y q son números primos impares distintos, entonces p es residuo cuadrático de q si y solamente si







LOS NUMEROS COMPLEJOS son de la forma $a+bi$ (donde a y b son números reales, e i es raíz cuadrada de -1). Pueden representarse mediante pares ordenados de números (a,b), esto es, como puntos de un plano, de la misma forma que los números reales pueden representarse mediante puntos de una recta. Los números complejos pueden manejarse así geométricamente. Por ejemplo, girar 90 grados una recta que una el origen con el punto (a,b) equivale a multiplicar el complejo (a,b) por i. (Se muestran en la figura tres giros de este tipo.) Gauss fue el primer matemático que observó que esta interpretación geométrica podía utilizarse para obtener definiciones puramente algebraicas de la adición y la multiplicación.

q es residuo cuadrático de p. Hay tan sólo una excepción a esta regla: si p y q son ambos de la forma $4n+3$, entonces uno es residuo cuadrático del otro, pero el segundo no lo es de éste. Este resultado es, a primera vista, de gran especialización; pero nos permite decir si un número primo impar es residuo cuadrático de otro número primo haciendo una pregunta que, con frecuencia, es más sencillo responder: ¿Es el segundo primo residuo cuadrático del primero? Este teorema ha inspirado algunas ideas nucleares del álgebra moderna y reviste gran importancia en toda la teoría de números y en otras ramas de la matemática. Tal era el valor que el propio Gauss le concedía que a lo largo de su vida llegaría a demostrarlo de ocho formas distintas.

Las *Disquisitiones* mostraron una tendencia que en Gauss se convertiría en modo de vida. Las demostraciones se pulen hasta relucir, suprimiéndose, si ello es posible, toda traza del proceso por el que han sido obtenidas, a fin de que tan sólo permanezca la estructura terminada. Gauss dijo en cierta ocasión: "Cuando se finaliza un noble edificio no deben quedar visibles los andamios". Las generaciones posteriores, que han tenido que afrontar el problema de comprender los métodos de Gauss, y no solamente sus resultados, pueden muy bien ser disculpadas de acusarle no sólo de haber retirado los andamios, sino de haber destruido los planos. En la investigación matemática, frecuentemente revisten mayor importancia las ideas y los métodos que los teoremas para los que fueron desarrollados. Una idea genuinamente buena puede ser generalizada a nuevos campos y dar en ellos los frutos que no pudieron preverse por adelantado. Dos aspectos de la matemática se encuentran aquí en conflicto: las matemáticas como forma del arte y las matemáticas como disciplina viva.

Este punto de vista no es enteramente moderno. Karl Gustav Jacobi, contemporáneo de Gauss, dijo de él: "Sus demostraciones son rígidas, heladas... lo primero que hay que hacer es descongelarlas". Otro contemporáneo, Niels Henrik Abel, observó: "Es como el zorro, que borra con la cola sus huellas de la arena".

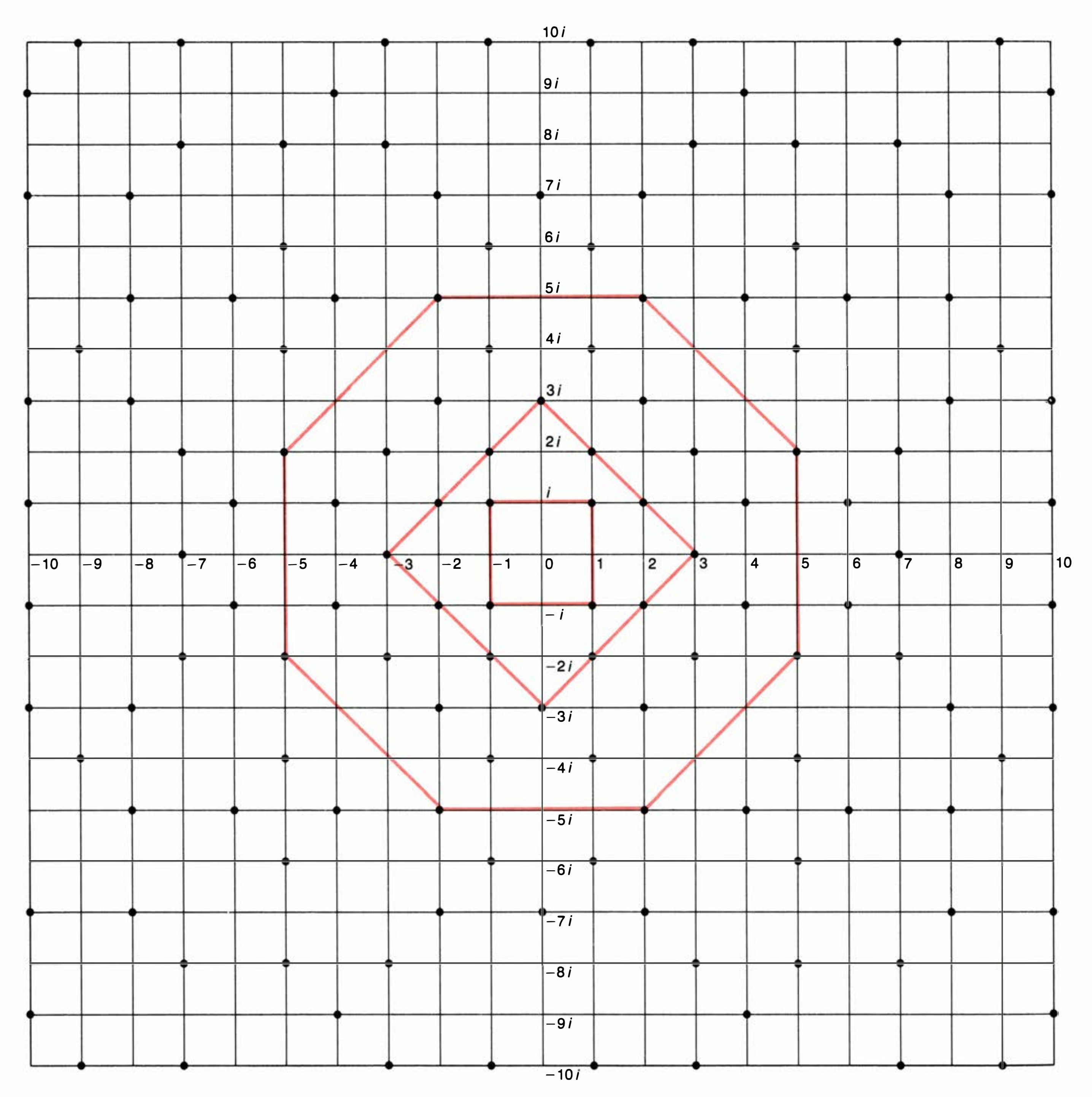
¿Por qué quiso Gauss ocultar sus métodos, prefiriendo dar solamente una síntesis y suprimiendo el análisis? Asumió el lema *Pauca sed matura* (Pocos, pero maduros), reflejo de su insatisfacción con los incompletos teoremas de sus colegas. Es imposible saberlo con certeza, pero quizá la penuria de su infancia le hizo circunspecto al ofrecer sus ideas. Tal vez pudo ocurrir también que no quisiera exponer trabajos incompletos por temor a verse ridiculizado en caso de error. Este temor es frecuente entre grandes matemáticos; a Newton, por ejemplo, hubo que persuadirle de que publicase sus *Principia.* Puede que esta actitud esté justificada; véase el caso de Georg Cantor, cuya obra sobre teoría de conjuntos y números transfinitos abre una nueva era de las matemáticas. Ridiculizado por sus más eminentes contemporáneos, está amarga experiencia fue, al parecer, causa de su locura.

Gran parte de lo mejor de la obra de Gauss en teoría de números estuvo relacionado con el problema de los números complejos, que se definen como números de la forma $a+bi$, donde a y b son números reales e i es raíz cuadrada de -1 (es decir, $i^2=-1$). Los números complejos fueron introducidos por los algebristas del Renacimiento, quienes les asignaron en generosa proporción propiedades místicas y descripciones caprichosas, como "real" e "imaginario". Incluso un hombre tan inteligente como Leibniz estuvo terriblemente confundido en este tema. Leibniz produjo auténticas antologías de disparates al tratar estos números, que no son ni positivos ni negativos. "El Divino Creador", escribió, "ha encontrado ocasión de manifestar su sublime inteligencia en esta maravilla del análisis, este portento del mundo ideal, este anfibio entre el ser y el no-ser que llamamos raíz imaginaria de la unidad negativa."

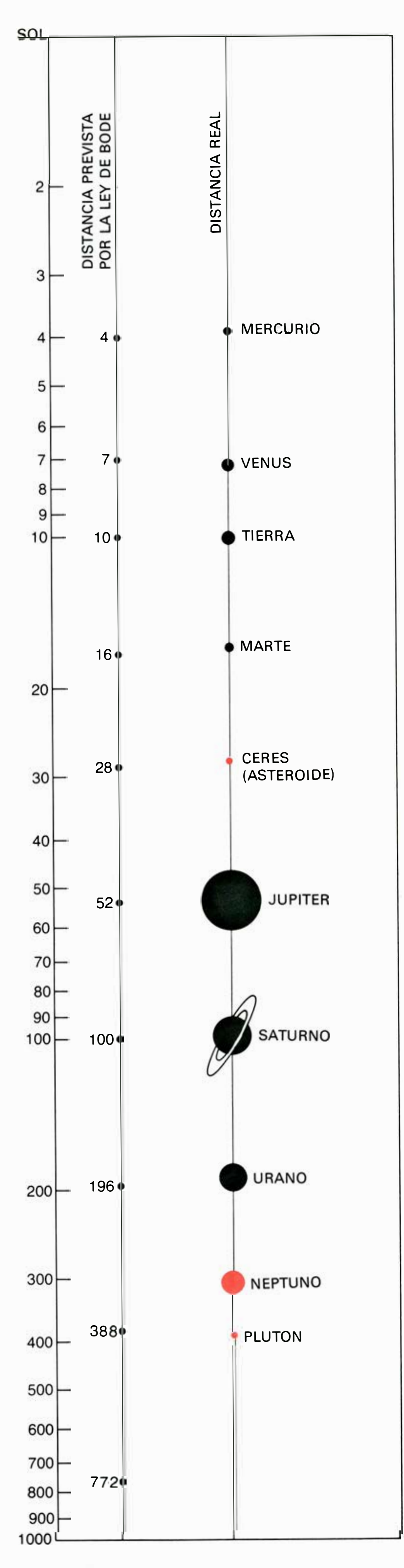
Gauss fue mucho más prosaico, y prefirió representar geométricamente los números complejos mediante puntos de un plano. A pesar de que en 1797 se publicó ya una exposición de este tipo, debida a un agrimensor noruego llamado Capar Wessel, quien preparó una representación analítica de la geometría plana esencialmente equivalente a los números complejos, este descubrimiento no fue conocido hasta 1897. Un contable suizo, Jean Robert Argand, desarrolló una des-

cripción parecida en 1806, y todavía hoy se le reconoce en los libros de texto la paternidad de la representación geométrica de los complejos. La contribución de Gauss consistió en ir más allá de la definición puramente geométrica de los números complejos. En una carta escrita en 1837 dice que en 1831 había comprendido que era posible evitar toda interpretación geométrica usando pares ordenados de la forma (a,b) en lugar de $a + bi$, dando definiciones puramente algebraicas de la adición y la multiplicación. Mediante pares ordenados mostró que las operaciones aritméticas con números complejos están definidas por las reglas: $(a,b) + (c,d)$ es igual a $(a + c,\ b + d)$ y $(a,b)\ (c,d)$ es igual a $(ac-bd,\ ad + bc)$. Es fácil comprobar que el par $(a,0)$ se comporta exactamente igual que el número real a y que $(0,1)^2$ es igual a $(-1,0)$. Así pues, el par $(0,1)$ es el misterioso número i, raíz cuadrada de -1. No se comprendió el verdadero valor de esta exposición hasta que fue publicada por William Rowan Hamilton en 1837. Gauss fue el primer matemático que hizo amplio y libre uso de los números complejos y les dio aceptación plena como objetos matemáticos genuinos.

En su tesis doctoral, presentada en la Universidad de Helmstedt en 1799, Gauss dio la primera demostración del "teorema fundamental del Algebra" (que hoy es más lógico demostrar como teorema topológico), a saber, que toda ecuación polinómica tiene una raíz compleja. También en este caso creyó Gauss que el teorema era muy importante, y, también, a



LOS NUMEROS PRIMOS GAUSSIANOS son números complejos de la forma $a + bi$ que carecen de factores de este tipo; se encuentran irregularmente distribuidos en el plano complejo. Gauss descubrió tres clases de tales números: (1) $\pm 1 \pm i$, que forman los vértices de un cuadrado; (2) $\pm p$ y $\pm pi$ (donde p es un número primo real de la forma $4n + 3$) que forman un diamante; (3) $\pm a \pm bi$ y $\pm b \pm ai$, que forman un cuadrado truncado. Ese tipo de números complejos que constituyen los primos gaussianos aparecen siempre en una de estas disposiciones.



LAS DISTANCIAS de los planetas al Sol quedan aproximadamente descritas por la ley de Bode, cuyo verdadero descubridor fue Johann Titius, en 1766. (Una unidad es igual a 14,9 millones de kilómetros.) La laguna de la serie que se conocía en 1800 precipitó la búsqueda del "planeta ausente", que culminó con el descubrimiento de asteroide de Ceres en 1801. Gauss realizó la difícil tarea de calcular la órbita de Ceres a partir de los escasos datos disponibles. Ceres volvió a ser observado donde él predijo. (Los planetas descubiertos después de 1800 se han representado en color. Obsérvese que la ley de Bode presenta aproximaciones mediocres para los planetas situados más allá de Urano.)

lo largo de su vida llegó a demostrarlo de hasta cuatro maneras distintas. La tercera demostración es particularmente típica de su estilo impenetrable y su original mentalidad. A partir de una ecuación polinómica, Gauss construye una complicada expresión en forma de integral doble. Si el polinomio carece de raíces, la integral doble dará el mismo valor calculándola por integración reiterada respecto de una variable y luego de la otra que al invertir el orden de integración. Gauss demuestra que no es así, sino que los distintos órdenes de integración dan a la integral valores distintos. Por consiguiente, la hipótesis de inexistencia de raíces tiene que ser falsa, esto es, existe una raíz.

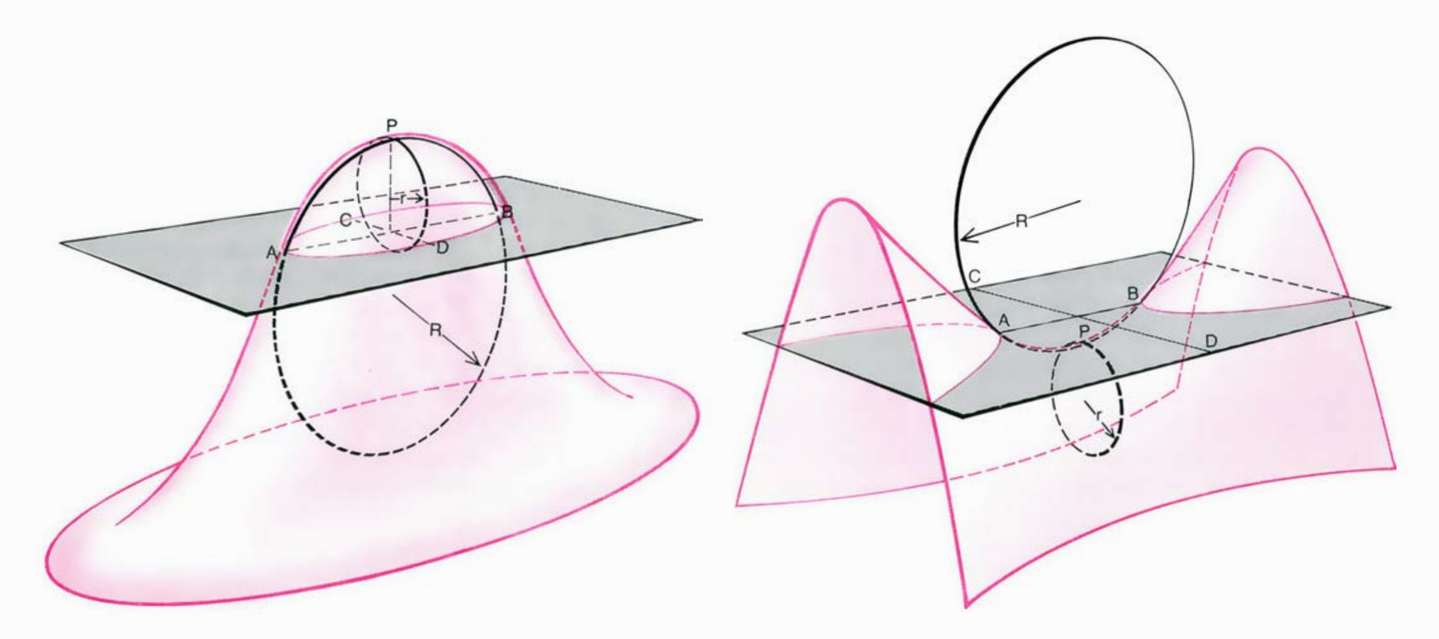
Para comprender de dónde vino la demostración hay que tener presente que Gauss poseía los teoremas fundamentales del análisis de variable compleja, pero no los había publicado. Su demostración es traducción de un razonamiento de teoría de funciones de variable compleja a técnicas ordinarias de variable real. Los resultados de la traducción son rigurosos desde el punto de vista lógico, pero hay una cierta perversidad en su formulación, que oscurece la idea inicial. Al parecer, Gauss creía que sus teoremas de variable compleja no estaban todavía suficientemente terminados como para publicarlos, y, por consiguiente, refundió convenientemente su razonamiento.

Gauss hizo mucho más con los números complejos. En 1811 descubrió el hoy llamado teorema de Cauchy: La integral de una función analítica compleja a lo largo de una curva cerrada que no rodee ninguna singularidad es igual a cero. Augustin Louis Cauchy convirtió este teorema en fundamento del análisis de variable compleja, y todavía hoy sigue siéndolo. No se le llama teorema de Gauss porque Gauss no llegó nunca a publicarlo. Parece verosímil que tuviera la intención de preparar una obra definitiva sobre análisis de variable compleja, pero que no tuvo nunca tiempo suficiente para desarrollar a su satisfacción.

Gauss elaboró también un método para descomponer números primos en producto de números complejos, obteniéndose algunos resultados dignos de mención. Por ejemplo, el número primo 2 puede descomponerse en la forma $(1+i)$ $(1-i)$. Así, 5 es expresable como $(2+i)$ $(2-i)$, 29 es $(5+2i)$ $(5-2i)$, etc. Sin embargo, ciertos números primos no pueden descomponerse, y permanecen primos (entre ellos, 7, 11 y 19). Gauss descubrió que aparte de 2, que es caso especial, los únicos números primos así descomponibles con números complejos son los de la forma $4n+1$, y que en tales casos la descomposición factorial es única. Más tarde se usaron métodos de este tipo para resolver problemas que en apariencia no tenían relación alguna con los números complejos.

En particular, Gauss utilizó números complejos de la forma $a+bi$, con a y b números enteros (llamados en la actualidad enteros de Gauss), para enunciar y demostrar una versión de la ley de reciprocidad cuadrática para residuos bicuadráticos. Se dice que el número k es residuo bicuadrático de otro número m si k es congruente módulo m a la cuarta potencia de un entero. Así, los residuos bicuadráticos de 10 son 0, 1, 5 y 6. La ley de reciprocidad bicuadrática enuncia que para dos números primos p y q existen ciertas relaciones entre los enunciados "p es residuo bicuadrático de q" y "q es residuo bicuadrático de p", con un cúmulo de condiciones relativas a p y a q. Este teorema es análogo al de reciprocidad cuadrática, pero es mucho más fatigoso de enunciar matemáticamente (y, por consiguiente, es muy difícil conjeturarlo, y, no digamos ya, demostrarlo). Si el teorema se generaliza al caso de ser p y q enteros de Gauss de la forma $a+bi$, pueden simplificarse notablemente tanto el enunciado del teorema como su demostración. Así, el paso a números complejos hace el problema más sencillo, y, su resolución, más natural que en el caso puramente real.

La demostración de Gauss de la ley de reciprocidad cuadrática utiliza enteros de Gauss, y constituye un modelo arquetípico de resolución de problemas de teoría de números. Primero, el problema se generaliza a un dominio de números complejos convenientemente elegido, llamado un cuerpo numérico, en el cual el problema admite un análisis más natural; a continuación, el problema se resuelve en este dominio, volviéndose a los enteros ordinarios al final de la demostración. Este potente método abrió las puertas de la hoy denominada teoría algebraica de números.



LA TEORIA INTRINSECA DE SUPERFICIES, desarrollada por Gauss, permite calcular la curvatura de una superficie midiendo solamente la longitud de las curvas contenidas en esa superficie. La curvatura de una superficie convexa se halla del modo siguiente: Primero, usando un plano paralelo al plano tangente en el punto *P*, se rebana la superficie cercana a *P* a lo largo de una elipse *(color oscuro)*. Se trazan los ejes mayor y menor de esta elipse, señalados *AB* y *CD*, y se proyectan sobre la superficie, obteniéndose las curvas *APB* y *CPD*. Se hallan a continuación dos circunferencias ("osculatrices") que tengan en el punto *P* el máximo grado de contacto con las curvas proyección. Sean sus radios *R* y *r*. Entonces, conforme el plano secante se aproxima a *P* y la elipse se contrae a este mismo punto, la curvatura tenderá hacia el valor 1/*Rr*. En *b*, la superficie presenta una ensilladura, y el plano la corta a lo largo de dos hipérbolas. Se conviene en que uno de los radios (y, por tanto, la curvatura) es negativo. En el caso de un plano, uno y otro radio, *R* y *r*, son de longitud infinita, y la curvatura es cero.

En 1801, Gauss se interesó por la astronomía, con lo que su trabajo matemático cambió bruscamente de dirección. A ello contribuyó, sin duda, su gusto por el cálculo. En toda su obra, incluso la más pura y erudita, hay largos cálculos; algunos de sus teoremas más profundos de la teoría de números fueron inducidos del examen de largas series de cifras. Además, Gauss solía proseguir muchos de sus cálculos hasta 21 cifras decimales, y esto, mucho antes de aparecer ningún tipo de máquina de calcular.

El interés de Gauss por la astronomía puede hacerse arrancar de un descubrimiento de Johann Titius, quien formuló, en 1776, una regla empírica para las distancias entre el Sol y los planetas. Titius dio inicialmente la serie 0, 3, 6, 12, 24, 48 y 96, en la que cada término es doble de su antecesor; más tarde, sumó 4 a todos los términos, con lo que se obtienen 4, 7, 10, 16, 28, 52 y 100. Resultó que estos números eran muy aproximadamente proporcionales a las distancias desde el Sol a Mercurio, Venus, la Tierra, Marte, Júpiter y Saturno, con la salvedad de que no había ningún planeta a la distancia 28. Esta regla, conocida hoy como ley de Bode (debido a que Johann Bode se la apropió sin mencionar su primitivo autor), no fue hasta 1781 sino una curiosidad, año en que William Herschel descubrió Urano, a distancia aproximada de 196 unidades. Como el siguiente término de la sucesión de Titius-Bode sería $2(96) + 4 = 196$, los astrónomos centraron su interés en la laguna correspondiente a 28.

En la Nochevieja de 1800 a 1801, Giuseppe Piazzi descubrió lo que pensó sería el planeta que faltaba. Se trataba de Ceres, que ahora sabemos que es uno de los millares de pequeños cuerpos del cinturón de asteroides situado entre Marte y Júpiter. Una vez divisado Ceres era importante calcular la órbita elíptica del nuevo cuerpo celeste antes de que los observadores lo perdieran de vista. La dificultad de observar un cuerpo tan pequeño hacía que los datos disponibles fueran escasos y poco fiables, y tanto más difícil la exacta determinación de su órbita. El propio Newton había hecho notar cuán difícil era la determinación de órbitas a partir de escasos datos. Para Gauss era la oportunidad de seguir las huellas del hombre más admirado.

Con tan sólo tres observaciones, Gauss preparó una técnica de cálculo de componentes orbitales tan precisa que a finales de 1801 y comienzos de 1802 varios astrónomos pudieron localizar Ceres sin ninguna dificultad. Parte de su técnica consistió en mostrar cómo las variaciones inherentes a la información de origen experimental podían representarse mediante una curva acampanada (muy conocida hoy con el nombre de distribución de Gauss). También diseñó el método de mínimos cuadrados, mediante el cual el valor estimado óptimo se determina haciendo mínima una suma de cuadrados de diferencias con los valores particulares de una serie de observaciones. Sus métodos, expuestos en 1809 en un artículo titulado "Teoría del movimiento de los cuerpos celestes que giran alrededor del Sol según secciones cónicas", son válidos todavía hoy. Tan sólo han sido necesarias unas cuantas modificaciones para adaptarlos a los modernos computadores.

Gauss tuvo parecido éxito en la determinación de la órbita del asteroide Pallas, para el que refinó sus cálculos con el fin de tomar en cuenta las perturbaciones de la órbita creadas por los otros planetas del sistema solar. En 1807 llegó a profesor de astronomía del nuevo observatorio de la Universidad de Göttingen, donde permaneció el resto de su vida. Su primera esposa murió en 1809, poco después del nacimiento de su tercer hijo. Se casó por segunda vez y tuvo otros dos varones y una hija.

Hacia 1820, Gauss volvió su atención hacia la geodesia, que estudia la determinación de la forma y tamaño de la Tierra. Dedicó a esta cuestión gran parte de los ocho años siguientes, tanto en estudios teóricos como en trabajos de campo. En 1821 fue nombrado consejero científico de los gobiernos de Hannover y Dinamarca, que le encargaron un estudio geodésico de Hannover mediante técnicas de triangulación. A tal fin,

Gauss puso a punto el heliotropo, instrumento que refleja la luz del sol en dirección exactamente especificada, haciendo así posible alinear a grandes distancias los instrumentos topográficos.

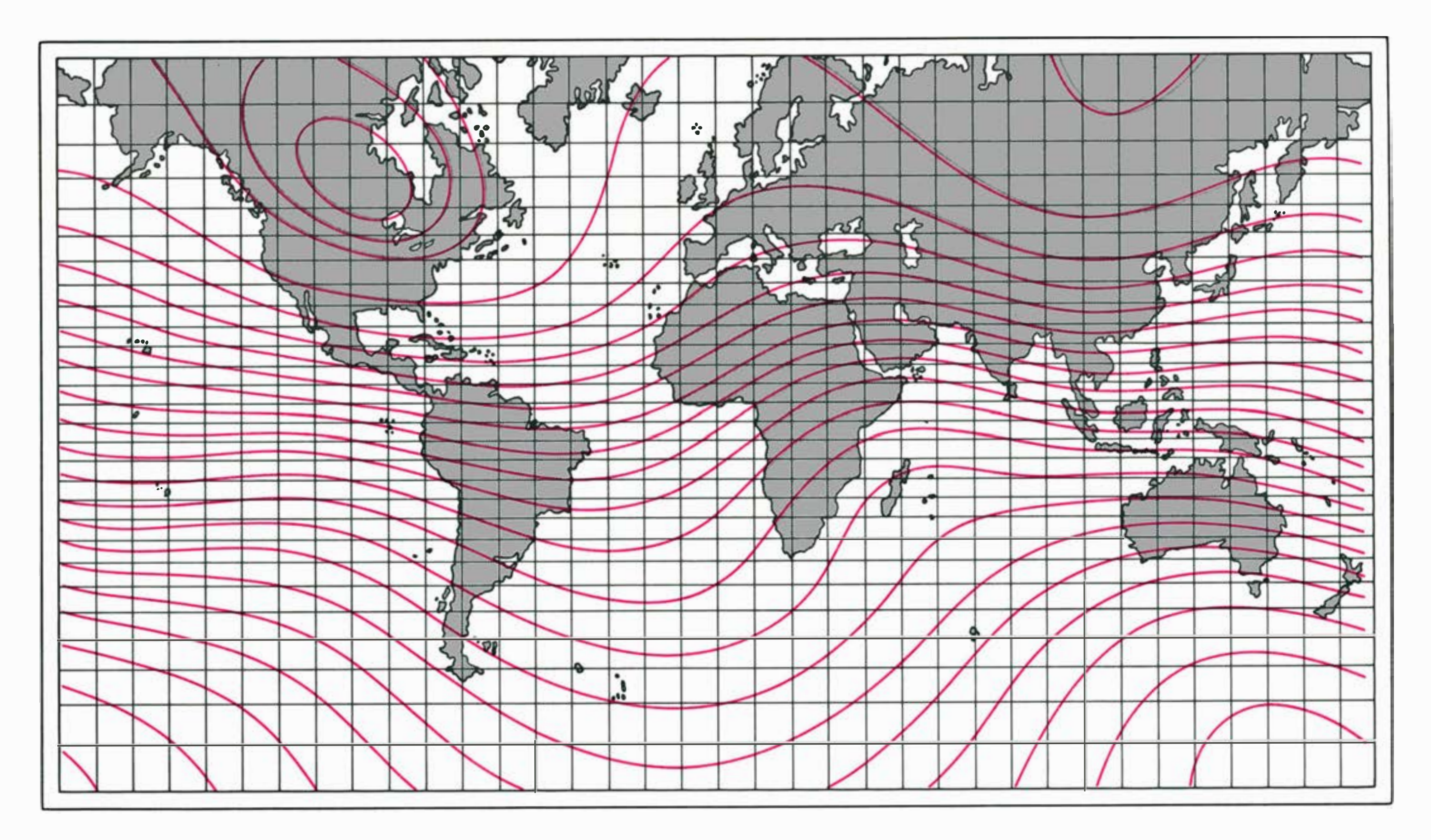
Los esfuerzos de Gauss para determinar la forma de la Tierra a partir de mediciones geodésicas reales le condujeron a la teoría pura. Trabajando con datos obtenidos en sus observaciones, desarrolló una teoría de superficies curvas mediante la cual las características de una superficie pueden determinarse con tan sólo medir la longitud de las curvas contenidas en ella. Esta teoría intrínseca de superficies inspiró a uno de sus discípulos, Bernhard Riemann, a desarrollar una geometría intrínseca general de espacios de tres o más dimensiones. Unos 60 años más tarde, las ideas de Riemann constituyeron la base matemática de la teoría general de la relatividad de Albert Einstein.

A partir de 1831, año en que el físico Wilhelm Weber llegó a Göttingen, Gauss trabajó en estrecha colaboración con él en la investigación, tanto teórica como experimental, del magnetismo. Ambos inventaron un magnetómetro, y como resultado de su común interés por el magnetismo terrestre, organizaron a través de toda Europa una red de observadores a fin de medir las variaciones del campo magnético terrestre. Gauss pudo demostrar teóricamente que el campo surgía del interior de la tierra, resultado de considerable importancia, porque delimitaba los posibles orígenes del campo, y hacía que la atención se concentrase en los mecanismos geofísicos que lo engendraban. Su contribución queda recogida en el "gauss", unidad de densidad de flujo magnético.

Gauss y Weber estuvieron también entre los primeros en señalar la posibilidad de transmitir mensajes por medio de la electricidad. Larga es la historia de la telegrafía, pero antes de 1800, o aledaños, tan sólo se usaban métodos no eléctricos. Hacia 1827 se transmitió un impulso eléctrico a más de 250 metros, lo que inmediatamente sugirió que la electricidad podría servir para la telegrafía. Se diseñaron diversos telégrafos eléctricos, pero ninguno llegó a ponerse a punto hasta 1832, año en que se conectaron los palacios de verano y de invierno del Zar, en San Petersburgo. Un año más tarde, Gauss y Weber disponían de un telégrafo que corría sobre los tejados de Göttingen, con longitud de 2,3 kilómetros. Las señales transmitidas consistían en una sucesión de cinco deflexiones de una aguja, bien a la derecha, bien a la izquierda (32 posibilidades en total); el dispositivo funcionó tan bien que los dos hombres lo utilizaban regularmente para comunicarse entre sí. El telégrafo de Gauss-Weber fue probablemente el primero que funcionó, en el sentido práctico del término, y se adelantó, en siete años, a la famosa patente de Morse.

La enorme fama de Gauss aumentó todavía más después de su muerte, al descubrirse, inéditos, numerosos resultados que anticipaban muchos de los principales progresos del siglo XIX. Además del teorema de Cauchy, había descubierto la doble periodicidad de las funciones elípticas, que, en manos de Abel y Jacobi, llegaron a ser el núcleo de la teoría de funciones del siglo XIX. Las funciones elípticas son funciones especiales $f(z)$ de una variable compleja z. Su doble periodicidad significa que hay dos constantes complejas distintas, que llamaremos a y b, tales que para todo valor de z, $f(z)$ es igual a $f(z+a)$, y también es igual a $f(z+b)$. La situación es análoga a las de las funciones trigonométricas, que tienen periodicidad simple. Así, por ejemplo, $\operatorname{sen}(z) = \operatorname{sen}(z + 2\pi)$. El descubrimiento de Gauss tuvo vigorosas implicaciones, debido a las numerosas conexiones entre la teoría de funciones de variable compleja y la teoría de números.

También estuvo Gauss entre los primeros que pusieron en tela de juicio que la geometría euclídea fuese la inherente a la naturaleza y al pensamiento humano. La geometría sintética de Euclides se ba-



MAPA MAGNETICO de la Tierra, basado en un dibujo publicado en 1840 en el *Atlas des Erdmagnetismus*, preparado conjuntamente por Gauss y Weber. Ambos obtuvieron las mediciones experimentales necesarias organizando una red de observadores por todo el mundo.

saba en ciertos axiomas, o proposiciones fundamentales, consideradas verdaderas por su propia evidencia. Todo el sistema geométrico se construía mediante razonamiento lógico sobre estos cimientos no demostrados. El axioma de las paralelas afirma que, por un punto exterior a una recta, solamente puede trazarse una paralela a ella. El axioma de las paralelas ha sido exhaustivamente estudiado a lo largo de toda la historia. No se impone al pensamiento con tanta fuerza como los restantes; además conlleva la noción de infinitud. Ya en la antigüedad, los matemáticos intentaron sustituirlo por axiomas más evidentes.

Este problema, todavía no resuelto en el siglo XVIII, recibió renovada atención, y muchos matemáticos y amateurs se esforzaron en demostrar que el axioma de las paralelas podía deducirse, lógicamente, de los restantes axiomas de Euclides. Todas las presuntas demostraciones que se obtuvieron resultaron contener falacias. Gauss tuvo noticia de la controversia siendo aún estudiante en Göttingen. En 1804 escribió una carta al matemático húngaro Farkas Bolyai donde le indicaba que la demostración que había dado Bolyai del axioma de las paralelas era falaz, porque había sustituido un razonamiento infinito por otro finito. Gauss incluyó en su refutación un comentario, donde expresaba haber tropezado ya en la misma dificultad. En 1815, no obstante, las recensiones que hizo de ciertos libros dejaban ver que en su opinión podría existir una geometría en la que no se verificase el axioma de las paralelas, y, a pesar de eso, fuese internamente coherente y libre de contradicciones. Dada la habitual circunspección de Gauss a la hora de manifestar sus propias ideas, las anteriores afirmaciones inducen fuertemente a pesar que él ya disponía con qué respaldarlas. Quizá porque su pensamiento iba contra la corriente de sus contemporáneos prefirió reservarlas; quizá creyera, probablemente con razón, que habrían de ser mal entendidas.

En 1820, el hijo de Bolyai, János, contagiado también de la fanática obsesión de su padre por demostrar el axioma de las paralelas, llegó a la conclusión de que tal demostración era imposible, y comenzó a desarrollar una nueva geometría que no utilizara el axioma de Euclides. Tres años más tarde había finalizado una memoria en la que proponía un sistema coherente de geometría no euclídea, que publicó como apéndice a un libro de su padre, *Ensayo sobre elementos de matemáticas para jóvenes estudiosos*. Gauss leyó esta memoria en 1832, y escribió a Bolyai padre que no podría ser elogioso con ella, porque hacerlo sería lo mismo que elogiar trabajos que él mismo había realizado 30 años antes. El joven Bolyai quedó muy decepcionado por la displicente respuesta de Gauss, y murió sin que casi nadie reconociese que había resuelto un problema enormemente importante, pendiente de solución durante muchísimo tiempo (problema que fue resuelto independiente y casi simultáneamente, de forma muy parecida, por Nikolai Ivanovich Lobachevsky). La actitud de Gauss fue muy injusta, en vista de que él mismo nunca tuvo la suficiente seguridad de su propio trabajo como para hacerlo público. Quizás estuviera algo celoso del éxito de Bolyai.

En muchos aspectos, Gauss se encontró en una encrucijada. Puede ser igualmente considerado como el primero de los matemáticos modernos que como el último de los grandes clásicos. La paradoja es fácil de resolver; sus métodos son de espíritu moderno, pero los problemas que afrontó fueron clásicos.

El marchamo de la obra de Gauss, muy especialmente en matemática pura, es su hábito de razonar con lo particular como si fuera general. Para usar con éxito esta técnica es necesario utilizar solamente aquellas propiedades del caso particular que tengan contrapartida general. Al afrontar un problema, Gauss combina la amplitud de los métodos generales con la intensidad y simplicidad de los casos particulares. Así, por ejemplo, su trabajo sobre los números complejos contiene en sí la simiente de la teoría general de los números algebraicos. Sus escritos dejan siempre en el lector la impresión de que Gauss sabía más de lo que decía, de que al explicar un resultado, estaba ya pensando en los problemas más generales que lo rodean, y que tenía ya idea de cómo empezar a resolverlos.

Podemos hacernos idea de la profun-

LA GEOMETRIA HIPERBOLICA es un sistema no euclídeo construido por Gauss. En él se puede hallar el área de un triángulo conociendo sus ángulos, lo que es falso en geometría euclídea. El diagrama aquí mostrado es parte de la demostración dada por Gauss de que el área del triángulo es proporcional a la diferencia entre 180 grados y la suma de los ángulos *A*, *B* y *C*. Debido a que el triángulo está dibujado en un plano euclídeo sus lados aparecen curvos, pero en el espacio hiperbólico son "rectos". Además, las líneas que aparecen agrupadas de tres en tres en los bordes del diagrama son todas ellas paralelas, con lo que la línea de apariencia curvada *XY* es paralela tanto a *XQ* como a *PY*, situación que resulta del todo imposible en geometría euclídea. Aunque todos los vértices del triángulo *XYZ* se encuentran en el infinito, su área es finita.

didad y fecundidad del pensamiento de Gauss observando algunas investigaciones recientes inspiradas por él. Por ejemplo, en 1947, André Weil, partiendo de ciertos teoremas de Gauss relativos al número de soluciones distintas y no congruentes módulo un número primo de las ecuaciones algebraicas, se vio conducido a formular tres conjeturas de gran trascendencia acerca de las variedades algebraicas sobre cuerpos finitos. Un cuerpo finito es un conjunto de elementos algebraicos, en número finito, que, además de reunir algunos otros requisitos, pueden sumarse, restarse, multiplicarse y dividirse entre sí, dando siempre resultados pertenecientes al mismo conjunto. Así, por ejemplo, las clases de restos de números enteros módulo uno primo p, forman un cuerpo finito.

Las conjeturas de Weil dan fórmulas que permiten calcular el número de soluciones de una ecuación algebraica en un cuerpo finito. Permiten, en particular, deducir si una ecuación tiene soluciones o no; información que, con las modificaciones necesarias, puede utilizarse para ecuaciones análogas en números enteros o números algebraicos. Como es obvio, las conjeturas de Weil son de carácter especializado, y están formuladas en lenguaje muy técnico. Han sido recientemente demostradas por Pierre Deligne.

Gauss enunció en alguno de sus trabajos la conjetura de que existe descomposición única en factores primos para números de la forma $p + q\sqrt{-D}$, donde D es un entero positivo, tan sólo cuando D sea uno de los números 1, 2, 3, 7, 11, 19, 43, 67 y 163. Este resultado, que él indujo por observación directa de series numéricas, ha sido demostrado recientemente por Harold M. Stark y Alan Baker, y conduce a nuevos e importantes resultados en teoría de números.

La importancia de Gauss es consecuencia de su capacidad para combinar lo general y lo específico. Forma un puente entre lo nuevo y lo viejo; sus ideas contienen la semilla de amplias teorías e importantes resultados. Es el más brillante caso de matemático capaz de extraer todo el jugo posible de un ejemplo maduro mediante razonamiento inductivo (pasando de los casos particulares a teorías generales) con preferencia al razonamiento deductivo (obtención de conclusiones específicas a partir de principios generales). En 1937, Eric Temple Bell escribió acerca de la influencia ejercida por Gauss sobre sus sucesores: "Vive en la totalidad de la Matemática". Afirmación que, en todo caso, es hoy aún más cierta que hace cuarenta años.

Juegos matemáticos

Las descomposiciones en partes iguales conducen hasta importantes regiones de las matemáticas

Martin Gardner

"Así pues, el Mago no perdió más tiempo, y adelantándose de un salto, alzó la afilada espada, hizo con ella uno o dos molinetes y de un potente tajo dividió exactamente en dos el cuerpo de la bruja.
Dorothy gritó..."
–FRANK L. BAUM: *El Mago de Oz*

Un tipo muy conocido de pasatiempos, de los que hay abundante provisión en libros antiguos, consiste en dividir una pieza de forma dada en dos, tres, o más, partes iguales. Unas veces, igual significa congruente; otras, es suficiente que las áreas de las regiones resultantes sean iguales. Los lectores habituales de esta sección podrán recordar muchos problemas de este tipo: enumerar de cuántos modos puede dividirse un tablero de ajedrez en dos, o en cuatro, piezas congruentes, cortándolo según la red de líneas que los cuadros delimitan; bisecar el símbolo ying-yang mediante una línea recta, a fin de obtener cuatro piezas de igual área; dividir un pastel cuadrado en n porciones de igual volumen; "reiterar losetas" descomponibles en varias copias a escala de sí mismas mutuamente congruentes; y otros muchos más.

Examinaremos hoy una variedad de problemas de división de piezas en partes iguales, no analizados previamente. Algunos de ellos llevan a importantes regiones de la moderna matemática.

Comenzaremos por el problema más sencillo: dividir una región plana en dos piezas congruentes. Aquí, congruentes significa que las dos piezas deben ser imagen especular una de la otra. Podría pensarse que todos los problemas de esta clase serán fáciles de resolver, pero en realidad pueden resultar de exasperante dificultad. Que yo sepa, no existe algoritmo alguno que permita decidir de modo general si una región de forma dada podrá o no descomponerse en dos o más piezas congruentes; los teoremas de interés a este respecto son de curiosa rareza.

Los lectores quedan invitados a medir sus fuerzas con las 12 formas de la ilustración de la página opuesta, intentando dividirlas en dos piezas idénticas. No hay ningún misterio. Cada mitad resultante es simplemente conexa (no tiene agujeros ni partes unidas por tan sólo un punto). Algunas figuras de éstas se han tomado de la sección de matemáticas recreativas que Pierre Berloquin escribe para una revista francesa. (Recientemente han sido publicadas por Scribner's las traducciones al inglés de tres populares libros de problemas de Berloquin.) La figura perforada es difícil, y abre nuevo campo en matemáticas recreativas: descomposiciones en partes congruentes de piezas con agujeros.

Obsérvese que la cuarta figura tiene solamente una simetría axial, cuyo eje la atraviesa verticalmente por su centro. Cualquier figura plana o tridimensional que posea simetría bilateral puede, evidentemente, descomponerse en dos mitades congruentes, cortándola a largo de su eje o su plano de simetría. En la ilustración de la página 110 se muestra la forma como el Mago de Oz taja en dos partes congruentes a un malvado Mangaboo, atentamente observado por Eureka, el gatito de Dorothy. (La escena transcurre en la Ciudad de Cristal, cuyos habitantes son hortalizas.) ¿Es necesario cortar siempre a lo largo de un eje o un plano de simetría para descomponer en dos partes congruentes una figura con simetría bilateral? La cuarta figura muestra que la respuesta puede ser negativa. El lector podrá, sin duda, hallar una descomposición de esta figura distinta de la trivial.

Por lo común, la dificultad de dividir una figura en n partes congruentes aumenta al crecer n, muy especialmente si se especifica de antemano la forma que han de tener las partes resultantes. Los 12 pentominos de la ilustración superior de la página 111 suscitan un interesante problema de descomposición en cuatro partes congruentes. (Es preciso mencionar aquí que "pentomino" es, desde 1975, marca registrada propiedad de Solomon W. Golomb, inventor del término.) ¿Cuántos de estos pentominos pueden descomponerse en cuatro piezas congruentes? Resultan serlo todos menos tres. Quedé sorprendido al descubrir que los nueve descomponibles pueden resolverse por repetición de un mismo módulo –un pentomino más pequeño– y que, además, este módulo es único. ¿Podrá descubrirlo el lector, e identificar los tres pentominos indescomponibles?

Eliminemos ahora el requisito de congruencia, exigiendo solamente que las n partes tengan iguales áreas. Parece intuitivamente obvio que cualquier forma plana podrá ser dividida mediante una línea recta en dos mitades de igual área. Quizá no sea tan evidente que esta línea puede ser paralela a cualquier recta dada, exterior a la figura. Para demostrarlo usaremos un famoso teorema: Si una función de una variable es continua en un intervalo cerrado de extremos A y B, la función toma en ese intervalo un valor máximo y un valor mínimo, así como todos los valores reales comprendidos entre ellos.

Veamos un ejemplo. Se está ascendiendo a una montaña por un sendero tortuoso, desde A hasta B. La altitud es, en cada momento, función continua de la posición en el sendero. El teorema nos dice que hay en el sendero al menos un punto de altitud mínima (pudiendo haber varios si el camino sube y baja), al menos un punto de altitud máxima, y al menos un punto en el que la altitud toma cualquier valor comprendido entre la máxima y la mínima anteriores. El teorema parece trivial de puro evidente, a pesar de lo cual, a la hora de demostrar teoremas no triviales, su potencia es fantástica.

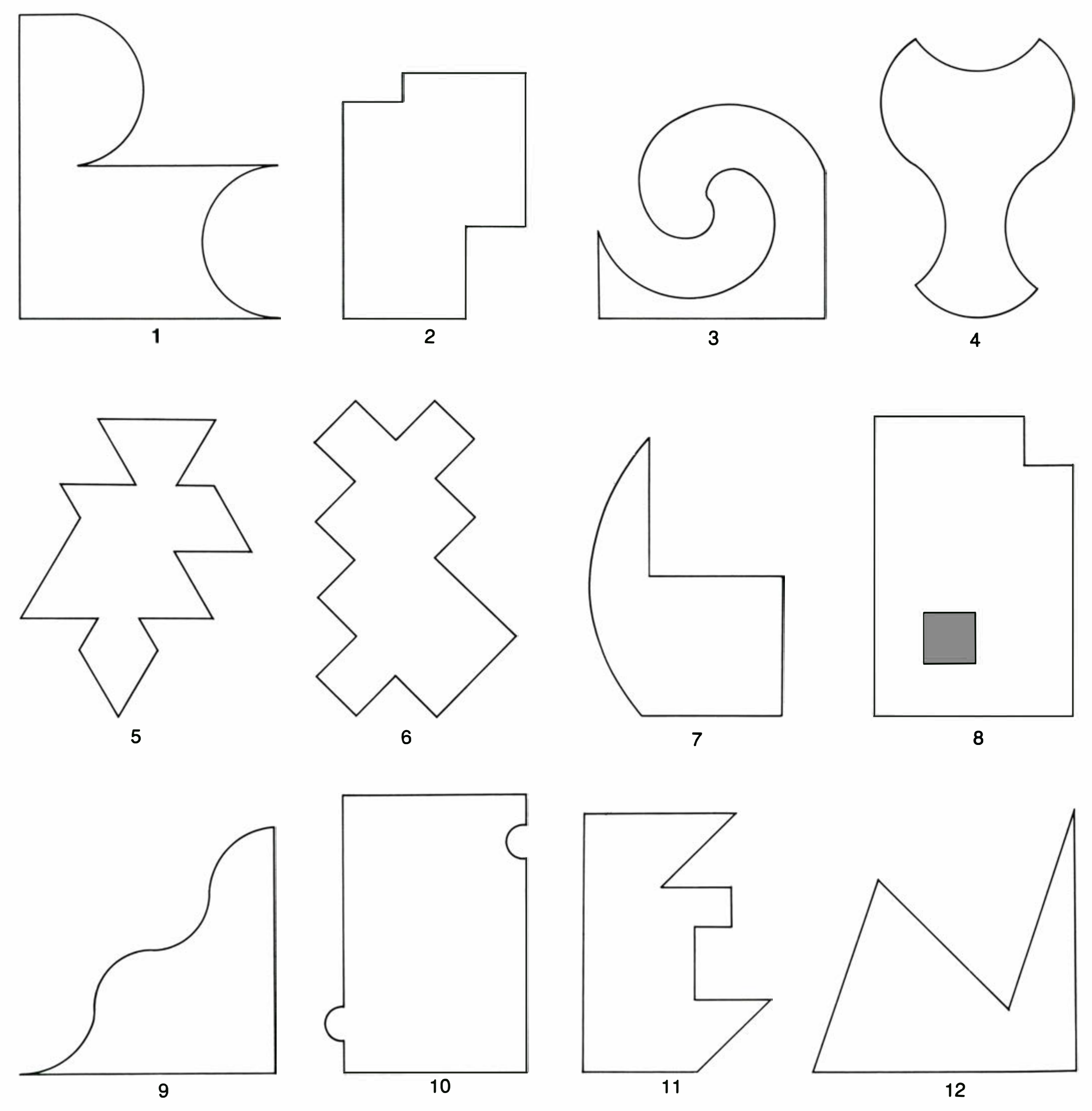
Fijémonos en la región sombreada del lado izquierdo de la ilustración central de la página 111. En su exterior tenemos una recta arbitraria x. Deseamos demostrar que puede trazarse una línea recta paralela a x que divida a la figura en dos regiones de áreas exactamente iguales. Imaginemos una recta desplazándose lentamente, siempre paralela a x, en la dirección perpendicular a x, señalada con flecha de trazos. La recta móvil toca a la región en A, y la abandona en B. Conforme la recta avanza, la región barrida es función continua de la distancia recorrida desde el punto A. El área situada del lado de A es nula en A.

y máxima en *B*. De acuerdo con nuestro teorema, en algún lugar intermedio el área es exactamente la mitad de la máxima. En esa posición, la recta biseca en dos la región dada.

La demostración es tan general que subsiste no solo para figuras conexas cualesquiera, incluso con agujeros, sino que también es válida para figuras disconexas. Una ojeada al lado derecho de la ilustración central de la página 111 debería bastar para convencernos de que siempre se puede trazar una recta a través de cualquier número de regiones, de forma tal que el área total a uno de los lados de la recta sea igual a la situada al otro.

Supongamos dadas dos regiones de formas arbitrarias en el plano. ¿Es siempre posible trazar una recta que divida simultáneamente a ambas regiones en sendas partes de áreas iguales? Podemos demostrar que sí. Primero se divide la región blanca *(véase la ilustración inferior de la página 111)* con una línea recta que no intercepte a la región gris. Giremos esta recta, de modo que siempre se tenga una bisección de la región blanca. Sabemos que es posible, por el teorema precedente. Como muestra la ilustración, la línea giratoria tocará en *A* a la región gris, y la abandonará en *B*. Conforme la recta va barriendo la figura, desde *A* hasta *B*, el área situada del lado de *A* crece continuamente desde cero hasta el máximo. Así pues, habrá una posición intermedia en la que el área del lado de *A* sea mitad del área total.

Acabamos de demostrar para el plano un famoso teorema que puede generalizarse a todos los espacios de dimensión superior. En el espacio tridimensional, pueden bisecarse por medio de un plano los volúmenes de tres cuerpos cualesquiera; en el espacio de cuatro dimen-



Piezas a descomponer en mitades congruentes

siones, los hipervolúmenes de cuatro figuras cuatridimensionales cualesquiera pueden bisecarse mediante un hiperplano de dimensión tres (N-1). Para espacios tridimensionales el teorema se denomina a veces "teorema del bocadillo", por ser aplicable a un bocadillo generalizado formado por dos rebanadas de pan y una de jamón, no importa cuáles sean las formas de estas piezas ni su posición en el espacio: siempre hay un plano que simultáneamente divide en dos partes iguales a cada una de las tres.

La generalización a espacios N-dimensionales exige matemática de alta potencia, pero, en dimensiones dos y tres, las demostraciones son consecuencias sencillas de nuestro teorema fundamental.

En cierto sentido, pues, estas demostraciones dan un método para hallar una recta bisectriz o un plano bisector, pero en la práctica su determinación no es cosa fácil. No basta con unir los baricentros (centros de gravedad) de las figuras, porque una recta, o un plano, que pase por el centro de gravedad de una figura plana, o de un cuerpo del espacio, no tiene por qué descomponerla en dos mitades de igual área o volumen.

Dada una figura simplemente conexa, aunque no necesariamente convexa, ¿existe siempre una línea recta que efectúe simultáneamente la bisección del área y del perímetro? Existe, en efecto, y la demostración es muy semejante a la precedente. Se traza una recta que divida en dos al perímetro, al que corta en P y en Q, como se muestra en la ilustración superior de la página 112. Si al prolongar esta recta también se biseca el área, el problema está resuelto.

Supongamos que la recta anterior no determine una bisección del área. Liguemos a la recta una flecha que señale hacia la región de menor superficie, y movamos los puntos P y Q a lo largo del contorno, preservando siempre la bisección del perímetro. Esta maniobra hará que la recta trazada por P y Q varíe continuamente de orientación. Al hacerlo así, variará también continuamente la razón de las áreas que determina en la región; después de que la recta haya girado 180 grados, volverá a ocupar la posición inicial, pero la flecha estará ahora apuntando hacia la mayor de las dos regiones.

Fijémonos en la diferencia obtenida al restar el área de la región señalada por la flecha de la otra. Esta diferencia es al principio positiva, y, al final, negativa; su valor es función continua del ángulo de la recta, y, por tanto, debe existir un valor del ángulo para el que la diferencia sea nula, y, las áreas, iguales. El teorema puede generalizarse a regiones de dimensión superior. Un cuerpo sólido puede ser cortado por un plano de forma tal que su volumen y el área de su superficie queden ambos divididos en dos; en general, un sólido del espacio N-dimensional puede ser bisecado por un hiperplano N-1 dimensional que divida en dos partes iguales su hipersuperficie.

Otro bello teorema, que tampoco es evidente, enuncia que cualquier región –que, como antes, no tiene por qué ser simplemente conexa, y ni siquiera, conexa– puede siempre descomponerse en cuatro partes de igual área mediante dos rectas perpendiculares. La demostración, que se ajusta a las ideas anteriores, es tan sutil como maravillosa.

Imaginemos que se tiene una hoja de papel de celofán, mayor que la región. Se divide la hoja en cuatro cuadrantes mediante dos rectas perpendiculares X e Y. El cuadrante superior izquierdo es gris, y, colorado, el superior derecho. Se sitúa la hoja sobre la región, con la línea horizontal por debajo de la región, y la vertical a la derecha de la misma. Se hace deslizar la hoja hacia arriba hasta que la horizontal X divida en dos partes iguales a la región; una vez conseguido esto, se la hace deslizar hacia la izquierda (la recta X desliza sobre sí misma) hasta que también Y divida en dos a la región. Se marca la región (no la hoja transparente) con A, B, C y D, para señalar sus cuatro regiones así obtenidas. La disposición resultante puede verse en la parte superior de la ilustración inferior de la página 112.

Por construcción, A más B es igual a C más D, y A más C es igual a B más D. Restándole la segunda igualdad a la primera, resulta $B-C = C-B$, o sea, $2B = 2C$. Por consiguiente, B es igual a C y A es igual a D.

Si A es igual a B, el problema ha quedado resuelto. Supongamos que no es así, y que la región B sea mayor que la



El Mago de Oz biseca un Mangaboo

A. Por tanto, el área de la parte gris de la región restada de la parte coloreada da diferencia positiva.

Giremos la hoja transparente 90 grados en sentido contrario al de las agujas del reloj, preservando siempre la doble bisección de la región por medio de *X* e *Y*. Su punto de intersección podrá ir de acá para allá, pero al completarse el giro, las dos rectas ocuparán necesariamente la posición primitiva, como se muestra en la parte inferior de la ilustración, con la única diferencia de que ahora se han intercambiado los papeles de *X* e *Y*.

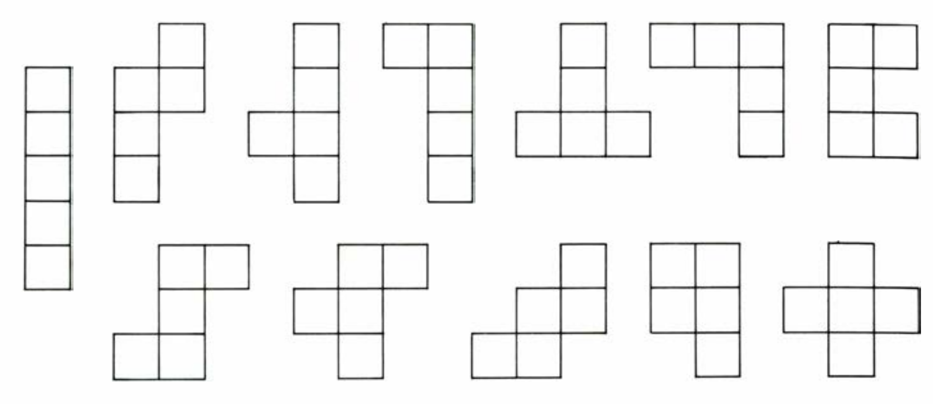
El área de la parte coloreada de la región es igual a *A*. El área sombreada de gris es igual a *C*, y como *C* es igual a *B*, el área gris es igual a *B*. Así que las áreas de las regiones gris y coloreada han intercambiado su posición. Por consiguiente, si se resta el área mayor, gris, del área menor, coloreada, se obtiene un número negativo.

Es fácil comprender cómo se aplica ahora el teorema fundamental. El valor obtenido al restar el área de la región gris del área de la región coloreada es función continua del ángulo girado al variar éste desde 0 hasta 90 grados. Puesto que el valor de la diferencia pasa de ser positivo a ser negativo, ha de existir un ángulo intermedio en el que la diferencia sea nula, y, las áreas, iguales. En esa posición intermedia, las rectas perpendiculares dividen exactamente en cuatro partes iguales a la región.

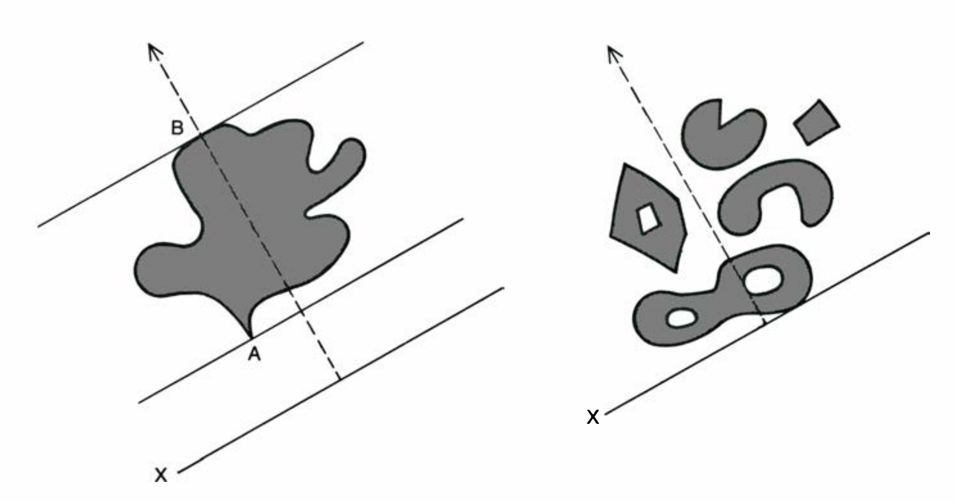
Este teorema se generaliza también a espacios de dimensión superior. Cualquier cuerpo sólido puede dividirse en ocho partes iguales mediante tres planos mutuamente perpendiculares. En general, cualquier sólido *N*-dimensional puede cortarse en 2^n partes de igual "volumen" mediante *n* "planos", mutuamente perpendiculares.

Al igual que en los ejemplos anteriores, el teorema no resulta excesivamente útil al afrontar prácticamente el problema de descomponer en cuatro partes figuras asimétricas. Sabemos, por ejemplo, que un triángulo de lados 3, 4 y 5 puede descomponerse en cuatro partes mediante dos rectas perpendiculares, pero la determinación efectiva de estas dos rectas es un problema completamente distinto. En realidad, no conozco ningún método sencillo de resolverlo.

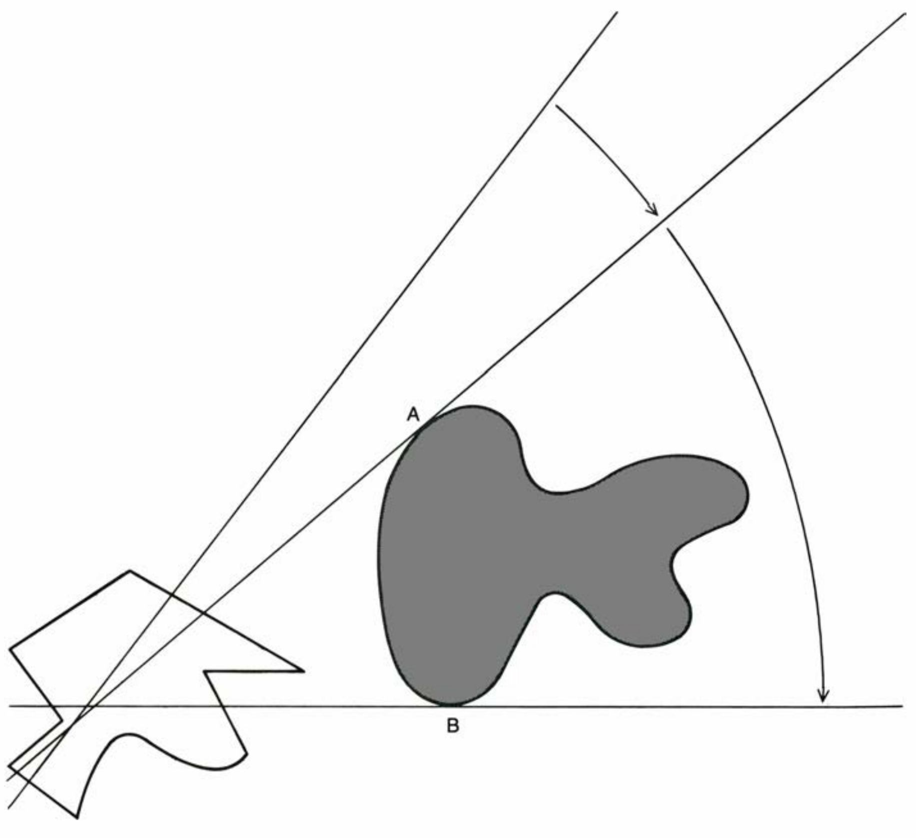
En la ilustración de la página 113 se ofrecen cuatro problemas de división en partes iguales para las que existen construcciones menos difíciles que la correspondiente al triángulo de lados 3, 4 y 5. Son, sin embargo, bastante artificiosas y exigen considerable ingenio.



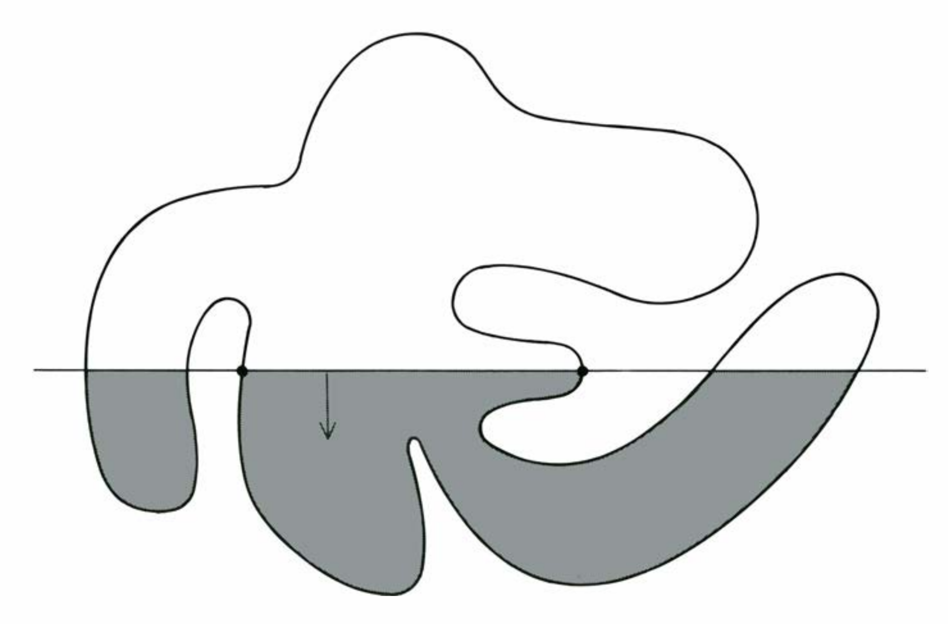
Los 12 poliominos de orden 5



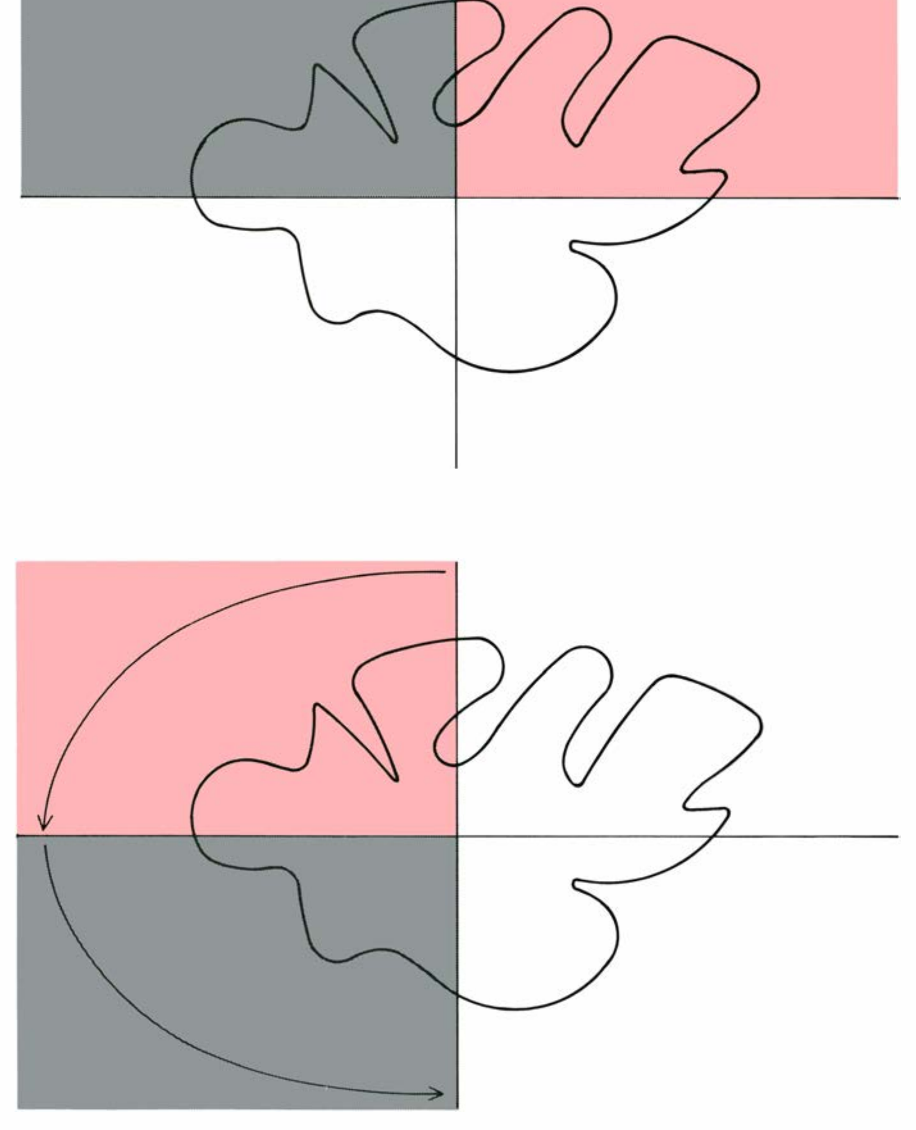
Demostración de un teorema de bisección de área



Demostración del teorema de bocadillo en el caso plano



Demostración de un teorema de bisección de área y perímetro



Dos perpendiculares descomponen en cuatro partes iguales una región

El primer problema consiste en dividir toda la región formada por nueve cuadrados en dos partes de igual área mediante una recta que pase por el vértice *P*.

El segundo problema es dividir en dos el área total de los cinco círculos trazando una recta que pase por el centro *P* del situado más a la izquierda.

El tercer problema consiste en dividir en tres partes de igual área un hexágono regular, mediante dos rectas que pasen por *P*.

El cuarto problema consiste en determinar una bisección de un triángulo equilátero mediante una línea de longitud mínima. La ilustración muestra una recta bisectora de longitud mínima; no obstante, existe una curva con esta misma propiedad que tiene menor longitud.

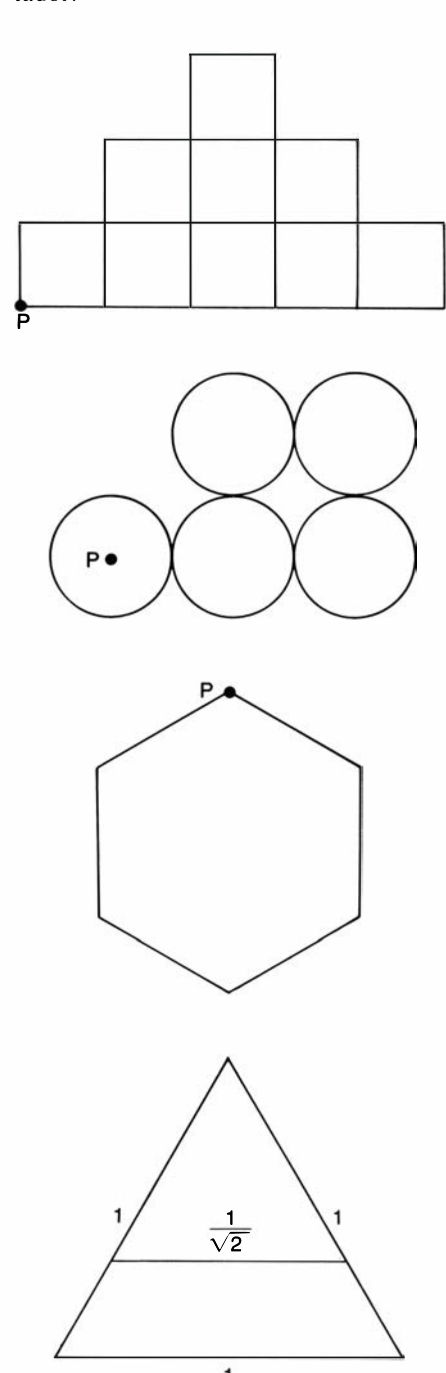
Se darán el mes próximo las soluciones de estos cuatro problemas, así como las de los anteriores.

El mes pasado presenté, junto con la solución de un problema sobre sombreros de distinto color, una nueva variante del mismo. Hay *n* personas, *n*-1 sombreros negros, y un solo sombrero rojo.

Las personas son "progresivamente ciegas", pues están sentadas en hilera, y cada una ve tan sólo los sombreros de las precedentes. *A*, situada al final, ve los sombreros de todos los demás, *B*, los demás menos el de *A*, y así sucesivamente. Se distribuyen al azar los *n* sombreros, y se pregunta a las personas (en el orden *A*, *B*, *C*...) si saben de qué color es su sombrero. ¿Podrá en todos los casos contestar que sí quien tenga el sombrero rojo?

La solución es curiosa. Si *A* tiene el sombrero rojo, es evidente que dirá que sí, pues ve *n*-1 sombreros negros. Si *B* tiene el sombrero rojo, dirá que no. ¿Por qué? Porque *A* tiene que contestar siempre que sí, y, por consiguiente, su "sí" no aporta ninguna información. Si *C* tiene sombrero rojo, dirá que sí, porque oirá decir sí tanto a *A* como a *B*. Si *D* llevara el sombrero rojo diría que no, porque el sí de *C* podría ser consecuencia de que *C* haya deducido que su sombrero, el de *C*, es rojo. El proceso se generaliza por inducción completa, llegándose a la conclusión de que el portador del sombrero rojo contestará afirmativamente si y solamente si ocupa posición impar desde el final. Cuando ocupe posición par, la contestación afirmativa del compañero situado inmediatamente detrás no le proporciona información suficiente para efectuar una deducción válida.

El examen de las curvas "monstruosas" realizado el pasado diciembre, curvas en copo de nieve, serpenteantes, y curvas que llenan un espacio, como las curvas de Peano, hizo uso del término "fractal", debido a Benoît B. Mandelbrot, para denominar las curvas cuya dimensión de Hausdorff era mayor que su dimensión topográfica. Acaba de ser publicado el libro *Fractals:Form, Chance and Dimension,* por W. H. Freeman y Cía. En él Mandelbrot escribe acerca de estas curvas y de sus muchas aplicaciones en la ciencia. Es un hermoso libro, profusamente ilustrado con fantásticas gráficas obtenidas mediante computador.



División de partes iguales

Taller y laboratorio

Cómo crear y observar una docena de arcos iris en una sola gota de agua

Jearl Walker

El arco iris, imponente y delicado al mismo tiempo, apareció en el cielo mientras me balanceaba con mi abuela en la mecedora de su porche. Aledo, Texas, es un pueblecito donde no hay mucho que hacer, salvo cosas parecidas a balancearse en el porche de la casa de mi abuela. A medida que el arco de color tomaba forma, ella se volvió a su nieto, el físico, y, con mirada socarrona, me preguntó por qué los colores del arco iris aparecen solamente en ese arco. ¿Por qué no se llena de colores todo el cielo?

Yo le di la explicación corriente, comparando la separación de colores en el arco iris con la dispersión de la luz blanca por un prisma. Se necesita una distribución geométrica especial de la dispersión para que cada uno de los colores alcance nuestros ojos, le expliqué. De aquí que los colores estén confinados a un arco único, un conjunto fijo de direcciones.

Satisfecho de mi explicación y sintiéndome bastante listo, me volví de nuevo hacia el cielo en el momento en que aparecía el segundo arco iris, algo más alto que el primero. Me bastó con volver a mirar a mi abuela para saber que tendría dificultades. Casi al unísono con nuestro balanceo, ella lanzaba pregunta tras pregunta. ¿Por qué el arco más alto es también el más ancho? ¿Por qué en el segundo arco está invertida la sucesión de colores respecto al primero? ¿Por qué el espacio entre los dos arcos está más oscuro que el cielo circundante? ¿Qué son esos arcos débiles y delgados que están debajo del primer arco iris y encima del segundo? Y, una vez más, si puede haber dos arcos, ¿por qué no más? ¿Por qué no está todo el cielo cubierto de arcos iris?

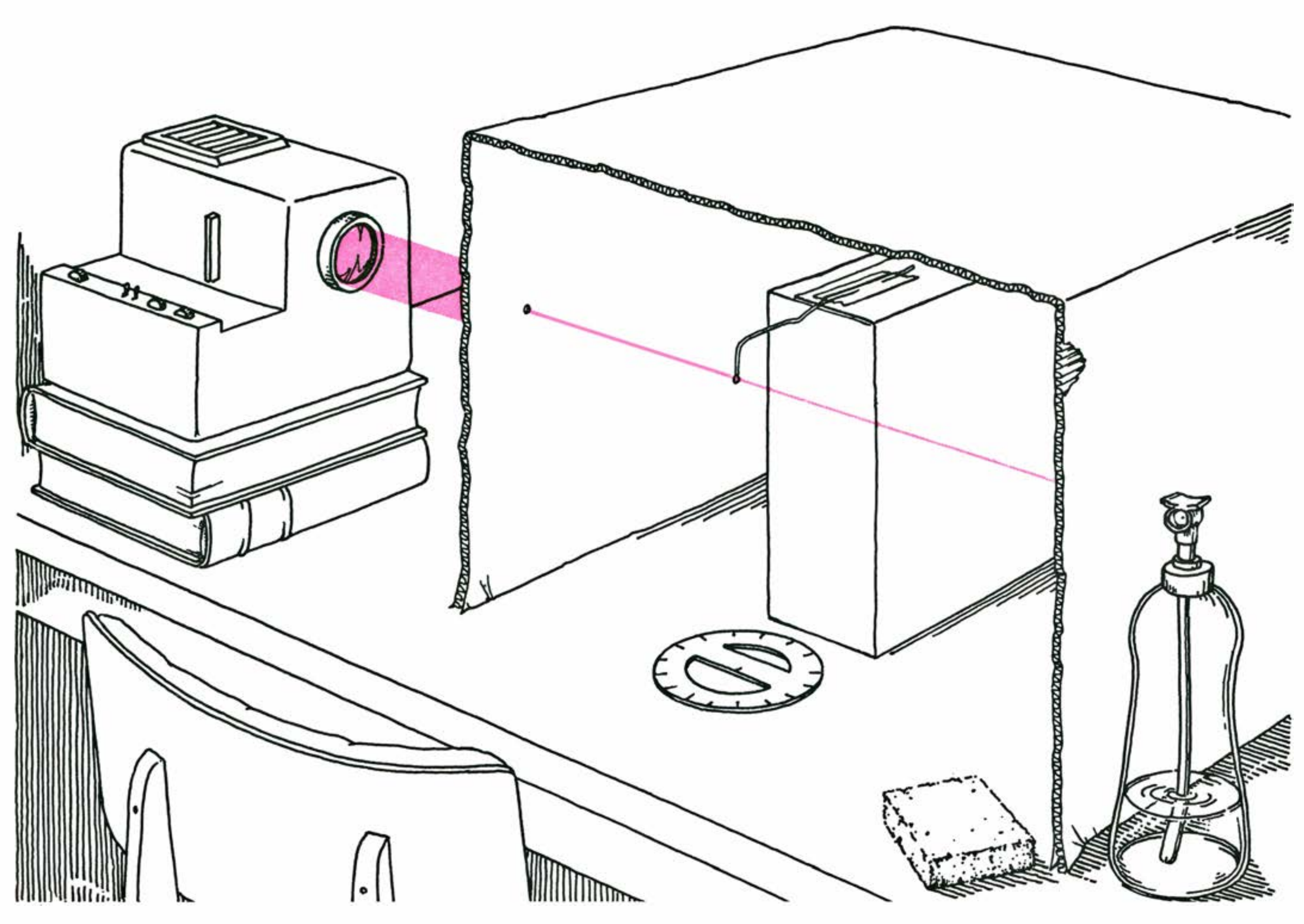
Cuando mi abuela empieza a preguntarme cómo está hecho el mundo, parece que mi base científica no logra dar alcance a su intuición. La pregunta de por qué el cielo no está cubierto de arcos iris es un buen ejemplo. En efecto, el cielo debería estar lleno de arcos iris, pero las observaciones de más de dos son raras. El arco iris primario, el que se ve más frecuentemente, está formado por rayos de luz que se han reflejado una vez en el interior de gotas de lluvia y han sido desviados hacia el observador. El arco iris secundario, que se ve algunas veces, requiere dos reflexiones internas de los rayos. Las reflexiones internas adicionales deberían dar lugar a arcos de orden superior, pero, al parecer, son demasiado débiles para que se puedan ver fácilmente sobre el resplandor del fondo. Están allí en el cielo, exhibiendo los mismos delicados colores del arco iris ordinario, pero no suelen verse.

H. Moysés Nussenzveig, de la Universidad de São Paulo, explicó la teoría del arco iris en el número de junio de *Investigación y Ciencia.* Dio cuenta de los intentos hechos desde el tiempo de Newton para explicar el aspecto del arco iris, culminando en recientes teorías matemáticas. Aquí me ocuparé principalmente de la observación de los arcos iris y, en particular, de un experimento mediante el cual se pueden ver muchos de los arcos de orden superior. Con cuidado, todos los arcos hasta el decimotercero deben ser observables. No aparecen en el cielo, sino en una simple gota de agua.

Muy pocas personas han visto arcos iris de orden superior. El investigador francés Felix Billet fue uno de ellos; en 1868 informó de haber observado 19 arcos en un delgado chorro de agua. En lugar de un chorro, nosotros emplearemos una gota suspendida de un fino alambre. El experimento es suficientemente sencillo para que se pueda realizar en la cocina y, sin embargo, permite que uno ingrese en ese pequeño club de personas que han visto arcos iris de orden superior.

Con un pulverizador o una regadera de chorro fino, pulverícese suavemente agua sobre el extremo de un trozo de alambre colocado verticalmente con su extremo libre hacia abajo. Continúese pulverizando hasta que se forme una gota y quede suspendida del extremo del alambre. Para obtener tal gota puede hacer falta práctica y paciencia. Puede

Rayo anaranjado del arco iris de primer orden



Aparato para observar arcos iris en una gota de agua

ser útil recubrir el alambre con cera negra, que ayuda a que el agua tome forma esférica y reduce también el resplandor al iluminar después la gota.

Una vez se ha formado una gota, cúbrase el aparato con una caja grande de la que se haya quitado uno de los lados. Hágase un agujerito en una pared de la caja adyacente al lado abierto y a la altura de la gota. La gota se ilumina a través del orificio y se observa por el lado abierto de la caja. La fuente de iluminación puede ser un proyector de diapositivas o de cine, que se debe colocar de manera tal que el haz de luz que penetra en la caja a través de orificio incida horizontalmente sobre la gota y la atraviese por el centro. Lo ideal es que sólo la gota se ilumine y no el alambre.

La principal razón por la que los arcos iris de orden superior no son visibles en el cielo es el resplandor de la luz solar transmitida directamente o reflejada por las gotas de lluvia. También en nuestro experimento tiende el resplandor a enmascarar los arcos iris, pero en la cocina se le controla más fácilmente que en la atmósfera. Solamente la luz que cae sobre una parte de la gota contribuye a un arco iris dado, y los puntos de fuerte reflejo se deben frecuentemente a luz que ha incidido en otra parte. Es, pues, posible eliminar o al menos reducir el resplandor interceptando una parte del haz incidente. Esto se realiza de forma muy sencilla ajustando una tarjeta de visita al orificio de modo que lo tape en parte.

Céntrese bajo la gota suspendida un transportador de círculo completo, con el cero hacia el lado de la gota opuesto al haz incidente. Con el transportador así alineado se puede medir el ángulo total de desviación de los rayos luminosos por la gota. Los rayos que pasan directamente emergen a cero grados; los reflejados de nuevo hacia la fuente de luz emergen a 180 grados. El arco iris de primer orden se puede hallar entre 137,6 grados (rojo) y 139,4 grados (azul), apareciendo los colores en la vecindad del borde derecho de la gota. El arco de segundo orden aparece hacia 129 grados al lado izquierdo de la gota. Cerca del centro se ven manchas luminosas de reflejo por ambos lados.

El arco iris de primer orden está formado por rayos que se han reflejado una sola vez en el interior de la gota, pero sólo algunos de esos rayos contribuyen a él; el resto emerge a ángulos inadecuados. Para simplificar las cosas, supongamos que sólo está iluminada una mitad de la gota de modo que los rayos incidentes se distribuyen uniformemente desde el centro a un borde, donde tocan la gota tangencialmente. Todos esos rayos son refractados cuando penetran en la gota; después son reflejados por la superficie interior y, finalmente, son refractados de nuevo al salir de la gota. La luz dispersada se reparte según un amplio intervalo de ángulos, viniendo determinada la desviación de un rayo dado por el punto donde incidió inicialmente sobre la superficie de la gota. Los rayos que inciden pasando por el mismo centro atraviesan la gota, se reflejan y son devueltos según la dirección de incidencia, sufriendo una desviación total de 180 grados. Los rayos que inciden más lejos del centro de la gota presentan menores ángulos de desviación, pero sólo hasta cierto punto. A medida que el rayo incidente se va acercando al borde de la gota, se alcanza una desviación mínima. El arco iris es simplemente el reforzamiento del brillo de la luz dispersada para el ángulo en que emergen mayor número de rayos.

El índice de refracción del agua, que determina la cuantía en que los rayos son desviados al penetrar en la gota y al abandonarla, es distinto para cada color, o longitud de onda, de la luz. (La luz azul es refractada más fuertemente que la roja.) Como resultado, cada color tiene un ángulo distinto de mínima desviación para el cual la luz dispersada presenta intensidad máxima. El arco iris subsistiría aun cuando el agua no dispersara la luz blanca en sus colores componentes, pero sería solamente una brillante banda blanca; la dispersión crea una serie de arcos monocromáticos, cada uno a un ángulo ligeramente diferente.

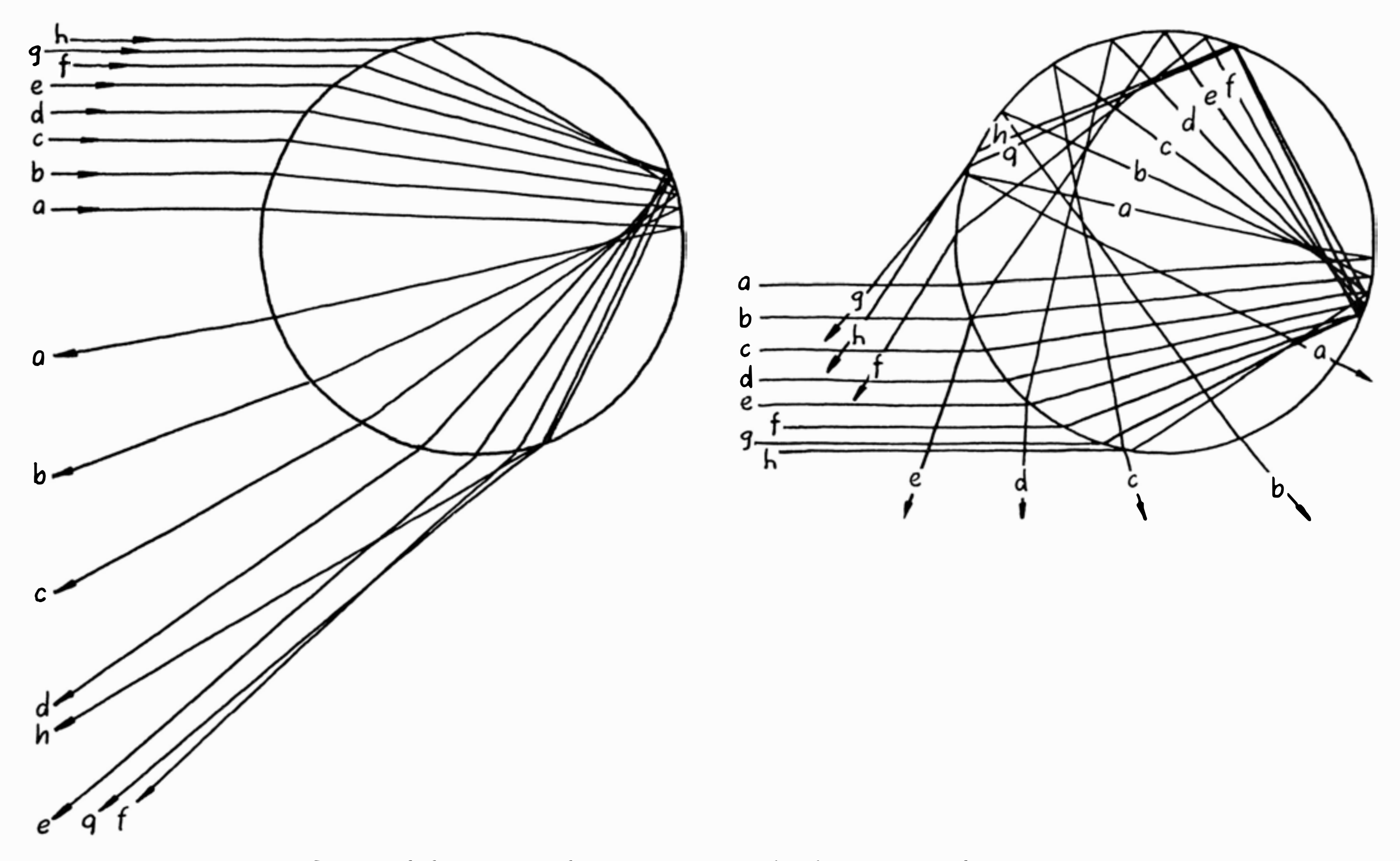
Incluso para el arco iris de primer orden formado en una gota de agua única, la mancha de luz blanca próxima a los colores del arco es comparativamente brillante. Esta mancha está formada por rayos reflejados por la superficie exterior de la gota. Si uno se encuentra a una distancia de la gota superior a un metro, como sin duda ocurre con las gotas de lluvia que forman un arco iris natural, la mancha reflejada y el arco iris serían indiscernibles, y el brillo efectivo de los colores del arco iris quedaría reducido. Los arcos iris naturales de orden superior suelen perderse por completo en el reflejo.

En la cocina, uno puede acercarse lo suficiente a la gota para distinguir los haces independientes de luz reflejada y de arco iris; además, se puede interceptar parte del reflejo. Muévase lentamente el borde vertical de una tarjeta cortando el haz incidente. Si la tarjeta se coloca por la parte del haz opuesta al observador, pronto bloquea los rayos incidentes que contribuyen al arco iris. Si se comienza desde el lado próximo, cuando la sombra del borde ha recorrido aproximadamente la cuarta parte del diámetro de la gota, intercepta los rayos que producen el reflejo. El bloqueo selectivo es más fácil con una gota grande, de unos dos milímetros de diámetro. A medida que la gota se evapora y se hace más pequeña, la colocación de la tarjeta resulta más difícil. La separación de los colores es también más clara con una gota grande, y se atenúa hasta el blanco a medida que la gota se evapora. Los arcos iris naturales con colores muy debilitados se observan cuando son gotas muy pequeñas las que dispersan la luz.

Los rayos que se reflejan dos veces en el interior de la gota, contribuyendo al arco iris de segundo orden, presentan también el máximo brillo para el ángulo de mínima desviación; una vez más, este ángulo es distinto para cada color. Si el arco iris secundario se observa desde el mismo lado de la gota que el primario, el azul cae a 126,5 grados y el rojo a 129,6. Como mi abuela indicó, la sucesión de colores es opuesta a la del arco iris primario, y el de segundo orden es más ancho. La anchura angular es mayor porque los rayos que dan lugar al arco iris de segundo orden penetran en la gota más próximos al borde y sufren una dispersión mayor.

Entre los arcos iris de primero y segundo orden, la gota dispersa relativamente poca luz. Los rayos que se han reflejado una vez dentro de la gota emergen todos según ángulos iguales o mayores que el ángulo del arco iris primario. Los rayos reflejados internamente dos veces emergen según ángulos iguales o menores que el del arco iris de segundo orden. La región angular entre ambos es, por tanto, relativamente oscura. La única luz que emerge en esa región es la que ha sido reflejada más de dos veces y, en consecuencia, es relativamente tenue. Este área oscura entre los dos primeros arcos se llama la "banda oscura de Alejandro", en honor del filósofo griego Alejandro de Afrodisias.

Todos los arcos iris de orden superior se forman mediante el mismo mecanismo que los dos primeros, pero tras reflexiones internas adicionales. Muchos de ellos son débiles y están enmascarados por el reflejo, pero se pueden encontrar si uno sabe dónde mirar. Examinando la gota a los ángulos calculados para los arcos iris de orden superior, yo pude observarlos todos hasta el de decimotercer orden. En varios casos hube de interceptar con una tarjeta los rayos inci-



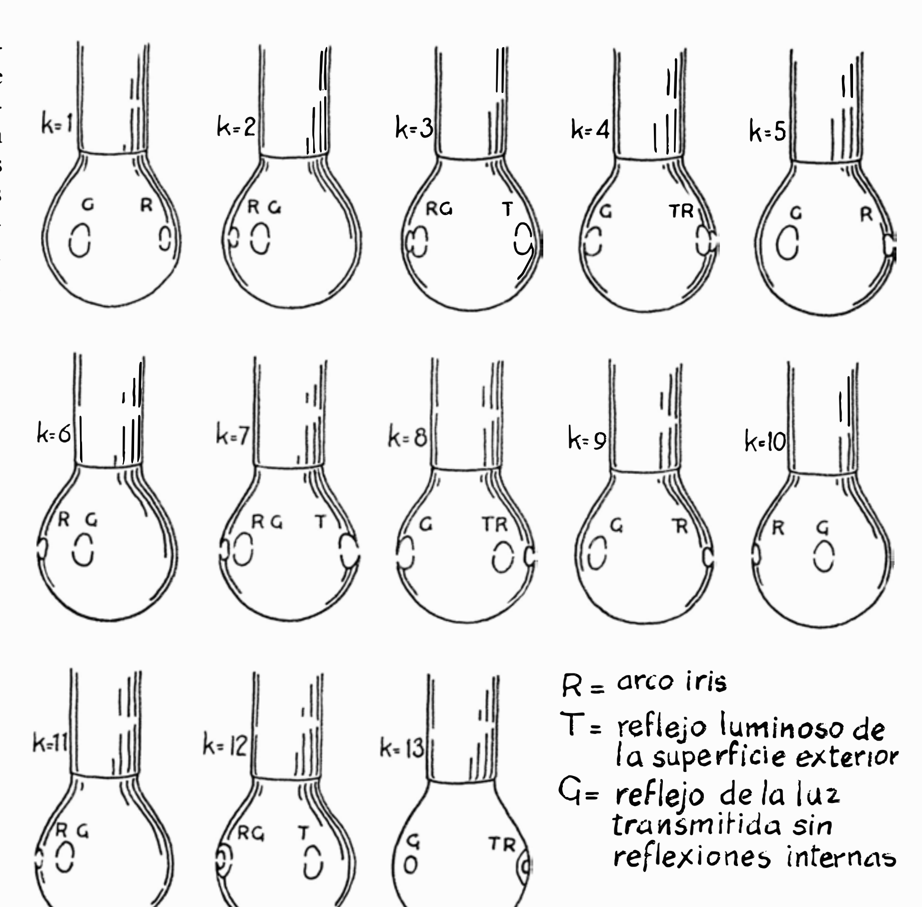
Caminos de los rayos en el arco iris primario (**izquierda**) *y secundario* (**derecha**)

dentes que contribuían al reflejo. Los arcos más difíciles de ver son los que se encuentran cerca del eje del haz incidente. Si uno tiene que mirar hacia la fuente de luz, como ocurre con los arcos cuarto, octavo y duodécimo, entonces no solamente hay que luchar con el reflejo producido por la superficie de la gota, sino también con la luz transmitida a través de ella sin reflexiones internas. Ambas son blancas y relativamente brillantes. Cuando se mira casi directamente según el haz, como en el caso del arco iris de décimo orden, es difícil evitar tapar el haz incidente con la cabeza.

El arco iris de undécimo orden es sorprendentemente acusado si se intercepta parte del haz incidente para eliminar el reflejo. Los arcos duodécimo y decimotercero requieren considerable paciencia para bloquear los suficientes rayos del reflejo, pero no demasiados del arco. Si Billet vio el arco iris de orden 19 en su aparato, o su paciencia o su vista eran mejores que las mías.

En la gota suspendida, puede aparecer luz coloreada mediante otro fenómeno además del arco iris. Una pequeña parte de la luz incidente es casi tangente a la gota y entra en ella casi paralelamente a la superficie. Para esos rayos, como para todos los otros, la luz azul se refracta ligeramente más que la roja. Si se observan los correspondientes rayos emergentes según el ángulo apropiado, sólo se verá la luz azul; la gota tendrá un borde azul. El ángulo correcto es aquel que resulta ligeramente grande para que sea alcanzado por los rayos rojos, pero que se halla dentro del intervalo de la luz azul, más fuertemente refractada. Los bordes azules son demasiado débiles para poder verse en la naturaleza porque es muy poca la luz incidente que entra tangencialmente en las gotas. Cada uno de los arcos iris tiene su propio ángulo de borde azul, pero, más allá de los dos primeros, los bordes son difíciles de encontrar. El borde azul de primer orden aparece al lado derecho de la gota a un ángulo de desviación de 165 grados.

Hay algunos ángulos cercanos a los correspondientes a los arcos iris en que aparecen manchas blancas que no son debidas a reflejo. Son rayos reflejados internamente que emergen según ángulos distintos de los de mínima desviación. Son blancas porque se componen de iguales intensidades aproximadamente de los distintos colores. Dos de tales manchas pueden verse sobre la gota desde un ángulo de 145 grados. Ambas resultan de luz reflejada una sola vez dentro de la gota, pero con diferentes puntos de incidencia, uno a cada lado del



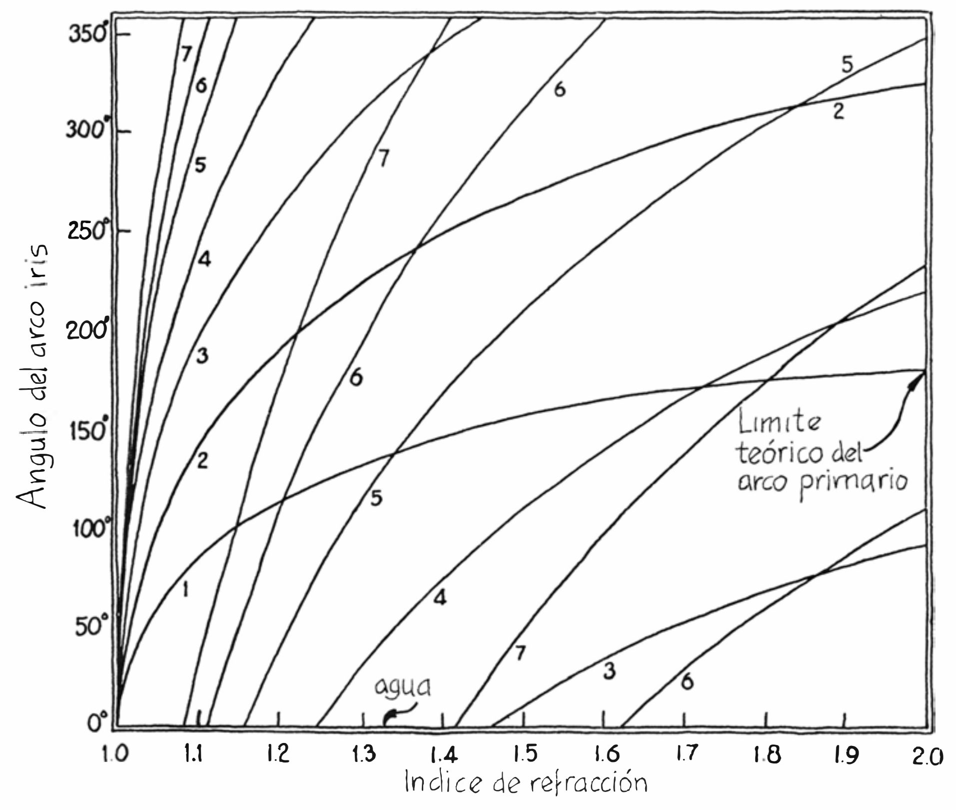
Aspecto de los trece primeros arcos iris

Orden del arco iris	Anchura angular	Rojo	Azul	Lado de la gota
1	1.72°	137.63°	139.35°	derecha
2	3.11	129.63	126.52	izquierda
3	4.37	42.48	38.11	izquierda
4	5.58	42.76	48.34	derecha
5	6.78	127.08	133.86	derecha
6	7.96	149.10	141.14	izquierda
7	9.14	65.59	56.45	izquierda
8	10.31	17.71	28.02	derecha
9	11.49	100.86	112.35	derecha
10	12.65	176.08	163.43	izquierda
11	13.82	93.11	79.29	izquierda
12	14.98	10.19	4.79	medio
13	16.15	72.68	88.83	derecha
14	17.31	155.51	172.82	derecha
15	18.48	121.69	103.21	izquierda
16	19.64	38.91	19.27	izquierda
17	20.81	43.84	64.65	derecha
18	21.97	126.58	148.55	derecha
19	23.13	150.70	127.57	izquierda
20	24.29	67.98	43.69	izquierda

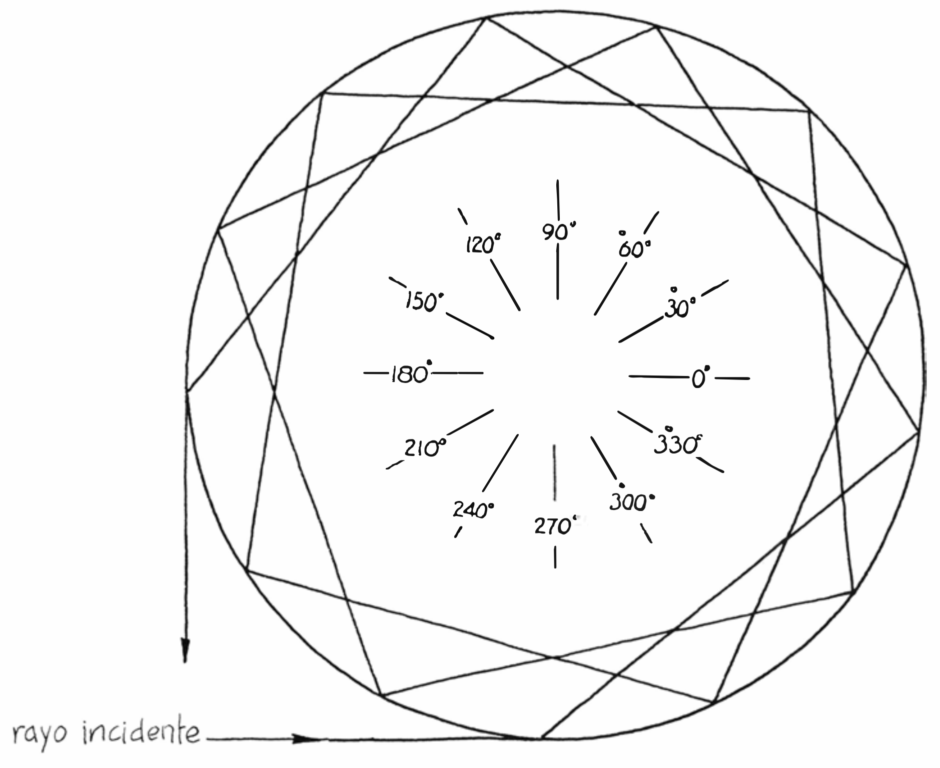
Ángulos calculados de los primeros 20 arcos iris

punto que origina el arco iris primario. Al variar el punto de vista desde un ángulo de 145 grados hacia el ángulo de 138 grados del arco iris, las dos manchas blancas se aproximan entre sí. Para el ángulo del arco iris, se reúnen y adquieren color.

El arco iris natural está polarizado paralelamente a su arco. Se puede determinar el sentido de polarización de los rayos de arco iris que emergen de la gota de agua suspendida observándolos a través de un filtro polarizador, como es un cristal de gafas polarizantes (que están proyectadas para excluir la luz con plano de polarización horizontal). La luz incidente puede considerarse como una mezcla a partes iguales de dos polarizaciones: una en el plano que pasa por la fuente de luz, la gota y el observador, y la otra perpendicular a tal plano. A través de la gota se transmite sin reflexión más luz con la segunda clase de polarización. El arco iris emergente está, por tanto, polarizado perpendicularmente al plano antes definido y paralelamente al arco.



Corrimiento de los arcos iris con el índice de refracción creciente



Formación del arco iris de undécimo orden

La posición angular de los arcos iris depende del índice de refracción del agua, que es aproximadamente 1,33 (para la longitud de onda de 5890 angstroms, que es la de la luz amarilla emitida por una lámpara de vapor de sodio). Yo busqué en la cocina de mi abuela algún otro líquido transparente que tuviese distinto índice de refracción. El más adecuado que encontré fue jarabe de maíz claro. Aplicando cuidadosamente pequeñas cantidades del jarabe al alambre, se puede fabricar una gota. Hay que trabajar de prisa, sin embargo, pues de lo contrario el jarabe se endurece y la gota se deforma. El jarabe fresco de maíz tiene un índice de refracción de 1,47 o 1,48. Como resultado, las posiciones angulares de todos los arcos iris giran en el sentido de las agujas del reloj respecto a sus posiciones en el agua, aunque no en cuantía uniforme. A medida que la gota se evapora a lo largo de los días y el jarabe se hace más concentrado, el índice de refracción aumenta. Los arcos iris giran lentamente en torno a la gota.

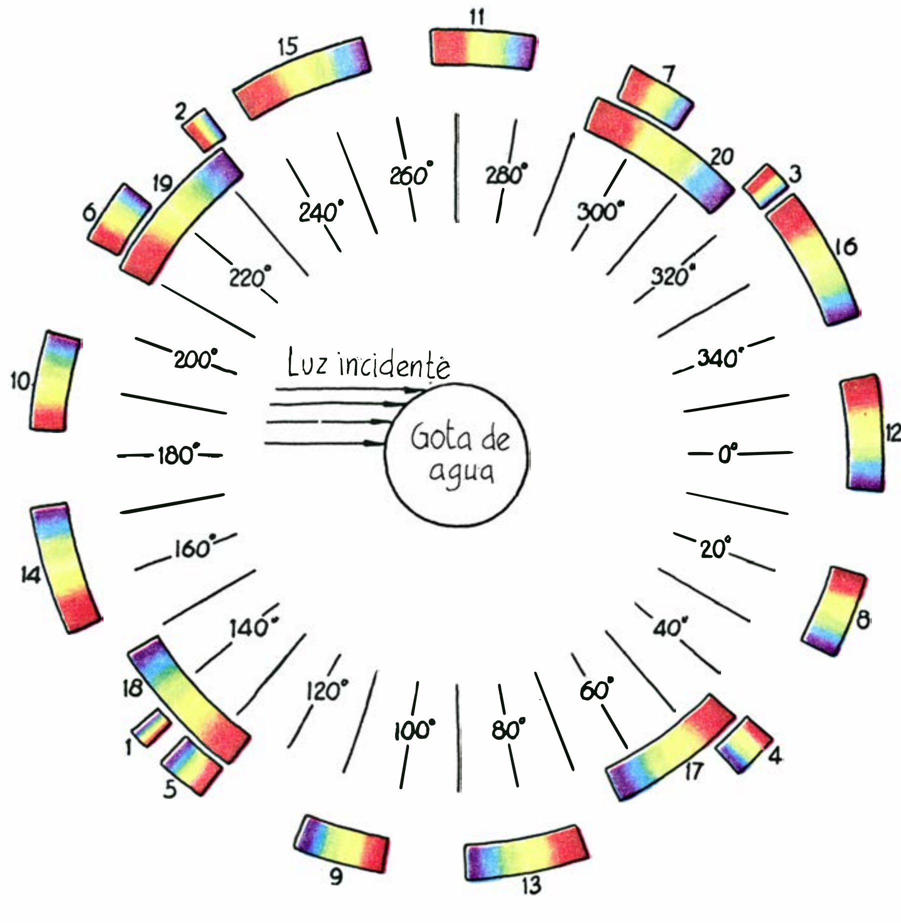
También he observado arcos iris formados en gotas de diiodometano, líquido transparente con el índice de refracción relativamente alto de 1,749 (en luz amarilla). El diiodometano puede obtenerse en establecimientos de productos químicos. Con el diiodometano no pude hallar los arcos iris de primero y cuarto orden porque están formados por rayos reflejados casi directamente hacia atrás en la dirección del haz incidente, y los de tercero y noveno orden quedaban enmascarados por el reflejo. Eran visibles el de segundo orden, los comprendidos entre el quinto y el octavo y el de décimo.

A medida que aumenta el índice de refracción desde el del agua hasta el del jarabe de maíz y el diiodometano, el arco iris de primer orden cambia hacia un ángulo de dispersión de 180 grados. Si el índice de refracción se hiciera exactamente igual a 2,0, el ángulo alcanzaría los 180 grados y el arco iris sería inobservable. Cualquier intento para verlo interceptaría el haz incidente. Aun cuan-

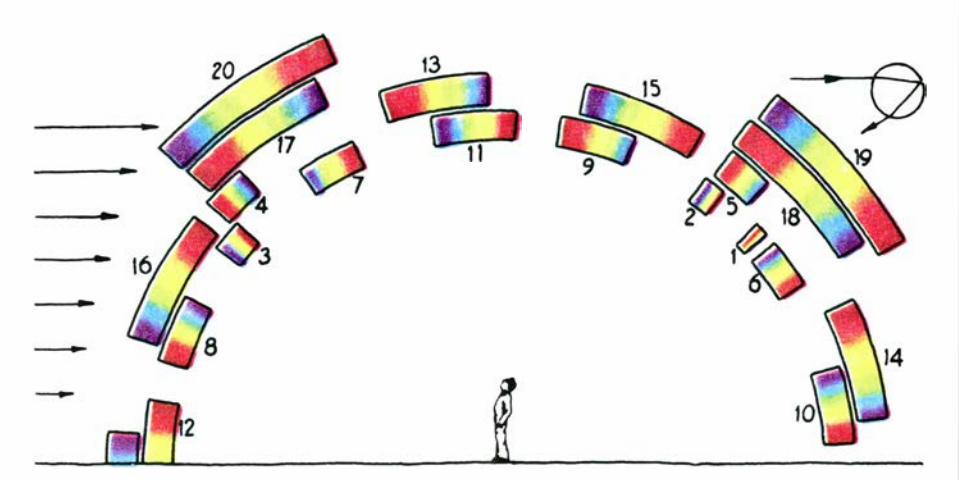
do el engorroso problema del ángulo de visión pudiera resolverse de algún modo, el arco iris de primer orden no aparecería. Con un índice de refracción de 2,0, todos los rayos que penetran en la gota se reflejan en el lado opuesto y son devueltos hacia la fuente luminosa. No hay dispersión y, por consiguiente, tampoco arco iris. Cada uno de los arcos iris de orden superior alcanza un límite análogo para un índice de refracción más elevado.

No todas las preguntas de mi abuela quedan contestadas con este experimento. Por ejemplo, no hemos explicado los débiles arcos que aparecen por debajo y por arriba de los arcos iris naturales, que se llaman arcos supernumerarios.

Tampoco lo ha sido el debilitamiento de los colores del arco iris a medida que la gota se evapora y se hace más pequeña. Pero, en fin, hemos visto los delicados colores de más de una docena de arcos iris dentro de una gota única de agua. ¡Qué fabuloso espectáculo sería que todos esos arcos pudieran contemplarse en el cielo azul!



Rosa de arcos iris rodeando una gota



Conjunto de arcos iris en el cielo

Libros

Redescubrimiento de la perspectiva lineal en el Renacimiento y el espejo mágico de Escher

Philip Morrison y Luis Alonso

THE RENAISSANCE REDISCOVERY OF LINEAR PERSPECTIVE, por Samuel Y. Edgerton, Jr., Basic Books, Inc., editores: THE MAGIC MIRROR OF M. C. ESCHER, por Bruno Ernst, Random House, Inc. Copérnico nació en 1473; su quinto centenario resonó por todo el mundo científico. Aquí celebramos, junto con un historiador del arte, experimental e imaginativo (de la Universidad de Boston), un momento mucho menos arbitrario del desarrollo del punto de vista renacentista. Cierto día del año 1425, "un hombre bajo, de edad madura, llegó a la plaza" situada entre la catedral de Florencia, aún inconclusa, orgullo de esa república y de sus ricos y generosos banqueros y mercaderes laneros, y el afamado baptisterio de mármol multicolor, enfrente del portal de la gran iglesia. Ese hombre era el escultor, arquitecto y artesano-ingeniero "Pippo di Ser Brunellescho el Florentino". Llevaba consigo una pequeña tabla pintada de unos treinta centímetros de lado, con un curioso orificio practicado en ella, y un espejo de igual tamaño. Se plantó dentro del portal, podemos suponer, mirando hacia fuera, hacia el baptisterio, manteniendo el cuadro en una mano "y lo que era bastante raro, levantado directamente contra su rostro". En la otra mano equilibraba el espejo, al que hacía entrar y salir de su visual cada cierto tiempo. "¿Qué es eso? ¿un reflejo especular en sus ojos o un destello de satisfacción?" Ingeniosamente, ha hecho coincidir su imagen pintada (tal como se ve mirando, a través del orificio del cuadro, su reflejo en el espejo) con lo que ve el ojo. Fue la primera demostración moderna de la perspectiva lineal con líneas que convergen en el punto de fuga. En efecto, el autor de un tratado de los años 1460 atribuía ese "sutil y hermoso" descubrimiento al hecho de que Brunelleschi "analizara lo que nos muestra un espejo". Tenía que ser un espejo plano, de vidrio liso, lo que en aquel entonces era escaso y costoso, motivo por el cual fue el espejo lo que determinó la escala de la tabla.

El profesor Edgerton lleva al lector a ese mismo portal y le muestra, con su propio espejo y cámara fotográfica y algunos planos, dibujos y fotografías persuasivos, cuál fue probablemente el campo visual, en dónde se situó el pintor para tomar su referencia del espejo, e incluso la leve inclinación que debe haber dado al espejo para equilibrar satisfactoriamente el cielo y el pavimento en su imagen. Otros eruditos han sugerido que Brunelleschi pintó directamente sobre el espejo (poco práctico e improbable) o que realizó el descubrimiento primero enrevesadamente, fundiendo las vistas en planta y laterales de la iglesia que sin duda estaba preparando. Fue por cierto el propio Brunelleschi quien cubrió el gran edificio inacabado con una de las mayores cúpulas catedralicias de Europa por esos mismos años.

El afortunado experimento de Edgerton constituye un poderoso argumento (más convincente de lo que él mismo parece creer) en favor de que Brunelleschi haya sido el primero en pintar con perspectiva, por haberla aprendido a partir de las imágenes formadas en espejos planos y por la expresiva manera de demostrar su descubrimiento a sus eruditos y pictóricos amigos mediante su impresionante artificio para superponer imágenes. Al mismo tiempo, con un margen de cosa de un año de adelanto o retraso, sus amigos, el pintor Masaccio y el escultor Donatello, crearon las primeras obras de arte en que claramente se emplea la perspectiva lineal. Su aristocrático amigo, el sabio Leon Battista Alberti, hombre singular por su educación científica y literaria y que había emprendido la pintura como solaz, escribió el primer tratado de perspectiva lineal como recurso para pintores.

Alberti preparó una versión de su obra en italiano (para que la leyeran los artistas) que dedicó, entre otros, a Brunelleschi, 10 años después del juego con espejos en la Piazza del Duomo. La maravillosa tabla pintada desapareció, pero el biógrafo de Brunelleschi afirmaba en los años 1480, "he tenido el cuadro en mi mano y lo he visto muchas veces", y describe el tamaño de la tabla, el orificio para mirar "pequeño como una lenteja en el lado pintado de la tabla", y el cuidado, delicadeza y exactitud de la obra. Por cielo, el pintor había puesto papel de plata bruñido, para reflejar las nubes naturales de modo que pudieran ser "movidas por el viento cuando sople".

Sólo un quinto de este librito vivaz y fresco está dedicado al relato del análisis y recreación del invento de Brunelleschi; en su mayor parte es un estudio de la historia de las ideas a partir de pinturas y textos importantes. Alberti es quien recibe mayor atención; una de las percepciones más claras se da dentro de su lenguaje, que muestra cuán central era la experiencia humana en su época. Se refiere a las superficies cóncavas como "las superficies interiores de cáscaras de huevo" y a un plano como compuesto de "muchas líneas... unidas y muy juntas entre sí como los hilos de una tela", haciendo vistosas metáforas del mundo de la experiencia cotidiana para reemplazar "el blanco y negro de Euclides".

Edgerton se cuida de no alabar en demasía a sus protagonistas; mediante el título mismo del libro asigna el primer lugar a los antiguos. Afirma que el propio Euclides, y también Ptolomeo (especialmente en las proyecciones descritas para cartografía en su *Geographia,* que apareció por primera vez en el Occidente cristiano hacia el 1400 en Florencia), habían poseído las ideas principales de Alberti y de sus contemporáneos.

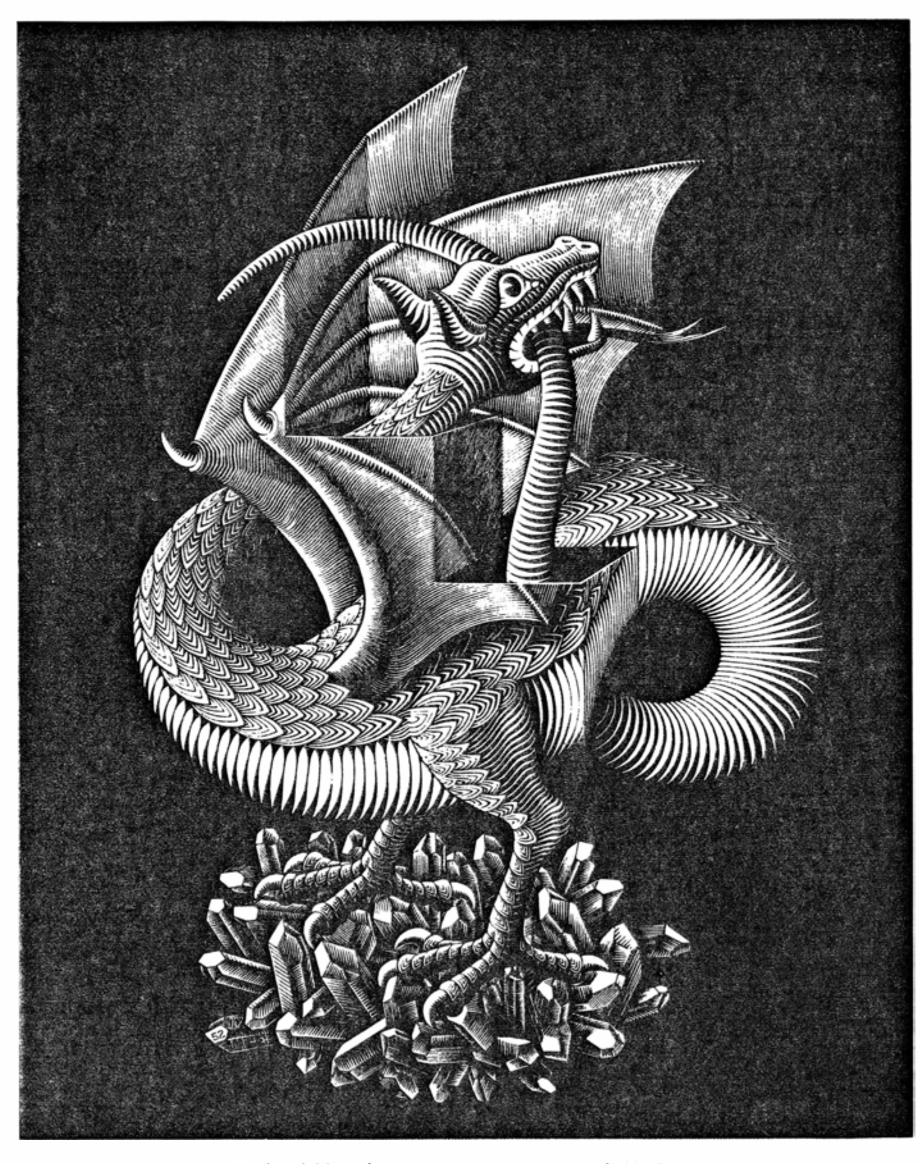
Lo que queda en claro es que la perspectiva lineal se ajusta a la grandiosa metafísica del Renacimiento. Fue su "forma simbólica", por emplear la expresión que Edgerton toma, a través de Erwin Panofsky, el célebre historiador del arte, del filósofo Ernst Cassirer. Con ella se construye un espacio en el cual, en determinado momento y posición, el observador se hace objetivo, un espacio basado en que sus experiencias hayan sido imparcialmente abstraídas. Lo que finalmente llegó a prosperar tras ese descubrimiento de parte de los artistas fue más ciencia que arte. La calidad íntegra de la perspectiva lineal, su espacio vacío libre de límite o de cambio, se acomoda a los espacios de Newton, a las lentes de Hooke y al cuaderno de viajes de Humboldt.

La ciencia ha madurado. Nuestros espacios proyectivos no excluyen a Newton, pero abarcan mucho más. Nuestro mundo es ensordecedor, variado, semicaótico, semiformado. Ya no basta con un punto de mira imparcial; el tiempo y el azar lo atraviesan todo. Los artistas han hecho también las paces con esta visión

del mundo, pero no la han guiado. Ha pasado ya largo tiempo desde que las percepciones de la ciencia fueran avanzadas por el pintor. Resulta admirable que un modesto grabador holandés llegase a expresar visualmente en sus meticulosos e ingeniosos grabados en madera y litografías buena parte de lo que nuestros osados y modernos postulados han agregado a la lucidez que fue captada por primera vez en la ciudad de Alberti.

Bruno Ernst fue amigo y vecino de M. C. Escher, y profesor de física y matemáticas y editor de una revista de matemáticas para estudiantes holandeses. Este libro excelente presenta muchos de los grabados más conocidos de Escher en el contexto de modelos, bosquejos, confidencias personales y las esperanzas que Escher expresó ante él y otros. ¿Puntos de fuga? Escher ciertamente usa la perspectiva lineal con singular habilidad. Pero he allí el famoso "Alto y Bajo", en el que hay dos puntos de fuga, uno que arrastra nuestra vista hacia los embaldosados cielorrasos, y el otro que la atrae hacia abajo, hacia las baldosas del piso. La relatividad ha hecho su entrada; el mismo cuadrángulo pequeño es a la vez cielorraso y piso, según donde se sitúe el observador. Se reproduce la litografía acabada, y con ella cuatro o cinco versiones bosquejadas y la explicación del propio Escher acerca de cómo un solo punto de fuga puede y debiera hacerse funcionar para dos observadores, superior e inferior.

También se presenta la notable retícula de Escher en la cual la escala aumenta constantemente conforme circundamos un punto situado en el centro de un dibujo. El espacio del grabado se infla enormemente: los maineles de las ventanas de la galería impresa no nos revelan un barco impreso sino un gran barco en la bahía. La escala se ha hecho explosiva; lo que Escher había hecho, según "los doctos caballeros... trataron una vez, en vano, de convencerme" era dibujar sobre una superficie de Riemann.



"Dragon" (1952), *de* **The Magic Mirror of M. C. Escher**

No tenía preparación matemática, pero llegó a los esquemas conceptuales de las matemáticas a través de su visión de artista, y de su empeño en presentar esa visión sin contradicciones. "Por lo que me atañía, se trataba tan sólo de una expansión o hinchazón cíclica, sin comienzo ni fin." Incluso la moderna teoría especializada puede hablar a un auditor perceptivo.

Los temas de Escher están puestos en orden y ejemplificados: diferentes estructuras de espacio que se penetran mutuamente: una superficie plana a la que se hace llenar, contener una aproximación a lo infinito dentro de un contorno limitado, para reflejar no menos el devenir que el ser; el conflicto entre los artificios ópticos de la perspectiva y los intrincados algoritmos del inconsciente mediante los cuales juzgamos la profundidad y la posición en un cuadro. Todo ello es un opulento legado, al que cada admirador de Escher acudirá para adquirir una visión tanto del hombre como de sus métodos. Los pequeños jinetes pasan cabalgando, fundiéndose uno en el otro; las últimas páginas nos muestran la última obra de Escher, el entrelazamiento de tres elegantes serpientes llenas de color retorciéndose doblemente en anillos sugiriendo la infinitud tanto en el centro como en el borde.

El libro empieza con un espejo mágico, el espejo de Escher, que deja a sus bestiezuelas en libertad de caminar en torno a una mesa embaldosada. Al volver la página vemos el punto final de la perspectiva lineal, el gran artesonado barroco, obra de ilusión óptica de Andrea Pozzo en Roma. La plenitud del mundo sigue todavía allí, no captada por el arte o por la ciencia. El libro muestra una instantánea en colores de Ernst con Escher unas semanas antes de la muerte del artista, en la que se lee una emotiva anotación: "estimo mi obra como la más hermosa y también la más horrible". También es notable en este caso cómo se puede asir en obras aparentemente convencionales de la época escolar la prefiguración de la madurez de un artista. (P. M.)

INICIACIÓN A LA FARMACOLOGÍA DEL COMPORTAMIENTO, de R. Bayés. Prólogo de F. G. Valdecasas, Editorial Fontanella, Barcelona 1977. Tras las obras de Thompson y Schuster e Iversen, ésta es, probablemente, la tercera que aparece en el mundo sobre la problemática psicofarmacológica enfocada bajo el prisma del denominado análisis experimental del comportamiento, y hemos de felicitarnos de que la misma haya sido escrita por un autor español.

Iniciación a la farmacología del comportamiento nos muestra la historia, posibilidades y técnicas de una nueva ciencia interdisciplinaria, la farmacología del comportamiento, fruto de la colaboración, a un mismo nivel de exigencia metodológica, entre farmacólogos y psicólogos, y cuyo objetivo es el estudio sistemático y acumulativo de los efectos de las drogas sobre el comportamiento de los organismos.

El ámbito del libro que comentamos, es, sin embargo, bastante más amplio. Aparte de explicar el nacimiento y desarrollo de esta joven disciplina –e incluir varios centenares de referencias bibliográficas como prueba de la seriedad y consistencia de sus logros científicos–, aporta una clara exposición de los siguientes aspectos: *a)* Importancia y consecuencias socioeconómicas de gran número de drogas, desde el alcohol y los medicamentos psicotropos hasta las drogas legalmente prohibidas. *b)* Establecimiento de las condiciones necesarias para que pueda establecerse un auténtico diálogo interdisciplinario entre psicólogos y farmacólogos. En este sentido, el planteamiento que ofrece Bayés puede considerarse como un posible paradigma para la colaboración entre unas disciplinas, como la psicología o la psiquiatría, cuyo valor científico ha sido puesto en duda numerosas veces, y otras de sólida categoría científica, como la farmacología, la fisiología, la bioquímica o la genética. *c)* Impugnación, a efectos de investigación científica, del modelo médico y de las clasificaciones clínico-terapéuticas ampliamente admitidas hasta el momento. *d)* Necesidad ineludible de trabajar con animales, aun cuando lo que se pretende en último término sea conocer los efectos de las drogas sobre el comportamiento humano en situaciones clínicas.

El libro de Bayés, a la vez que efectúa un análisis crítico de los puntos de vista sostenidos por la psiquiatría y la psicofarmacología tradicionales, y de la política investigadora que se lleva a cabo en este terreno, tanto en centros universitarios como en la mayoría de industrias farmacéuticas, propone una alternativa que, a pesar de que puede comportar algunos riesgos, se muestra sumamente atractiva.

Otro punto interesante abordado en la obra consiste en la exposición de algunas técnicas que se han empezado a usar recientemente para el estudio experimental de los factores ambientales responsables de las toxicomanías. (L. A.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

DEBATE SOBRE EL ADN RECOMBINANTE

MOLECULAR BIOLOGY OF THE GENE. J. D. Watson. Addison-Wesley Publishing Co., Inc., 1976.

NATIONAL INSTITUTES OF HEALTH GUIDELINES FOR RESEARCH INVOLVING RECOMBINANT DNA MOLECULES. Public Health Service. U.S. Department of Health, Education, and Welfare, junio 23, 1976.

RECOMBINANT DNA RESEARCH: BEYOND THE NIH GUIDELINES. Clifford Grobstein en *Science*, vol. 194, n.º 4270, págs. 1133-1135; diciembre 10, 1976.

NOTES OF A BIOLOGY-WATCHER: THE HAZARDS OF SCIENCE. Lewis Thomas en *The New England Journal of Medicine*, vol. 296, n.º 6, págs. 324-328; febrero 10, 1977.

RECOMBINANT DNA: FACT AND FICTION. Stanley N. Cohen en *Science*, vol. 195, n.º 4279, págs. 654-657; febrero 18, 1977.

LA ATMOSFERA DE MARTE

COMPOSITION AND STRUCTURE OF THE MARTIAN ATMOSPHERE: PRELIMINARY RESULTS FROM VIKING 1. A. O. Nier, W. B. Hanson, A. Seiff, M. B. McElroy, N. W. Spencer, R. J. Duckett, T. C. D. Knight y W. S. Cook en *Science*, vol. 193, n.º 4255, págs. 786-788; agosto 27, 1976.

ISOTOPIC COMPOSITION OF THE MARTIAN ATMOSPHERE. Alfred O. Nier, Michael B. McElroy y Yuk Ling Yung en *Science*, vol. 194, n.º 4260, págs. 68-70; octubre 1, 1976.

STRUCTURE OF MARS' ATMOSPHERE UP TO 100 KILOMETERS FROM THE ENTRY MEASUREMENTS OF VIKING 2. Alvin Seiff y Donn B. Kirk en *Science*, vol. 194, n.º 4271, págs. 1300-1302; diciembre 17, 1976.

HEPATITIS VIRICA

PROCEEDINGS OF A SYMPOSIUM ON VIRAL HEPATITIS en *The American Journal of the Medical Sciences*, vol. 270, n.º 1, julio-agosto, 1975; vol. 270, n.º 2, septiembre-octubre, 1975.

APPROACHING THE CONTROL OF VIRAL HEPATITIS TYPE B. Joseph L. Melnick, Gordon R. Dreesman y F. Blaine Hollinger en *The Journal of Infectious Diseases*, vol. 133, n.º 2, págs. 210-229; febrero, 1976.

ADVANCES IN VIRAL HEPATITIS: REPORT OF THE WORLD HEALTH ORGANIZATION EXPERT COMMITTEE ON VIRAL HEPATITIS. WHO Technical Report Series, n.º 602, 1977.

LA DEFENSA EN LOS OPISTOBRANQUIOS

COMPETENCIA I EVOLUCIÓ EN ESPECIES VEÏNES DE GASTERÒPODES MARINS. Joandomènec Ros en *Col·l. Soc. Cat. Biol., VII: Evolució:* 101-121, 1974.

SISTEMAS DE DEFENSA EN LOS OPISTOBRANQUIOS. Joandomènec Ros en *Oecologia aquatica*, 2:41-77, 1976.

BIOLOGY OF OPISTHOBRANCH MOLLUSCS, vol. I. T. E. Thompson, The Ray Society. Londres, 1976.

LA SOLIDIFICACION DEL CEMENTO

SOME STUDIES OF THE HYDRATION OF PORTLAND CEMENT USING HIGH VOLTAGE (1MV) ELECTRON MICROSCOPE. D. D. Double en *Materials Science and Engineering*, vol. 12, n.º 1, págs. 29-34; julio, 1973.

THE CONVERSION OF HIGH ALUMINA CEMENT/CONCRETE. C. Bradbury, P. M. Callaway y D. D. Double en *Materials Science and Engineering*, vol. 23, n.º 1, págs. 43-53; abril, 1976.

THE HYDRATION OF PORTLAND CEMENT. D. D. Double y A. Hellawell en *Nature*. vol. 261, n.º 5560, págs. 486-488; junio 10, 1976.

ANIONES DE LOS METALES ALCALINOS

ELECTRONS IN LIQUID AMMONIA. J. C. Thompson. Oxford University Press, Inc., 1976.

ELECTRON-SOLVENT AND ANION-SOLVENT INTERACTIONS. Dirigido por L. Kevan y B. C. Webster. Elsevier Scientific Publishing Co., Inc., 1976.

ALKALI METAL ANIONS: AN UNUSUAL OXIDATION STATE. James L. Dye en *Journal of Chemical Education*, vol. 54, n.º 6; junio, 1977.

EL OJO COMPUESTO DE LOS INSECTOS

THE SIZE OF OMMATIDIA IN APPOSITION EYES. Horace B. Barlow en *The Journal of Experimental Biology*, vol. 29, n.º 4, págs. 667-674; diciembre, 1952.

PHYSICAL ASPECTS OF THE SENSE ORGANS. Hl. de Vries en *Progress in Biophysics and Biophysical Chemistry*, vol. 6, pág. 208-264; 1956.

THE VISUAL SYSTEM OF INSECTS. Timothy H. Goldsmith y Gary D. Bernard en *The Physiology of Insects: Vol. II*, dirigido por Morris Rockstein. Academic Press, Inc., 1974.

THE COMPOUND EYE AND VISION OF INSECTS. Dirigido por G. Adrian Horridge. Oxford University Press, Inc., 1975.

PHOTORECEPTOR OPTICS. Dirigido por A. W. Snyder y R. Menzel. Springer-Verlag New York, Inc., 1975.

GAUSS

MEN OF MATHEMATICS. Eric T. Bell. Simon & Schuster, Inc., 1937.

DISQUISITIONES ARITHMETICAE. Carl Friedrich Gauss. Yale University Press, 1966.

CARL FRIEDRICH GAUSS: A BIOGRAPHY. Tord Hall. The MIT Press, 1970.

WERKE. Carl Friedrich Gauss. G. Olms Verlag, Hildesheim-New York, 1973.

JUEGOS MATEMATICOS

WHAT IS MATHEMATICS? Richard Courant y Herbert Robbins. Oxford University Press, Inc., 1941.

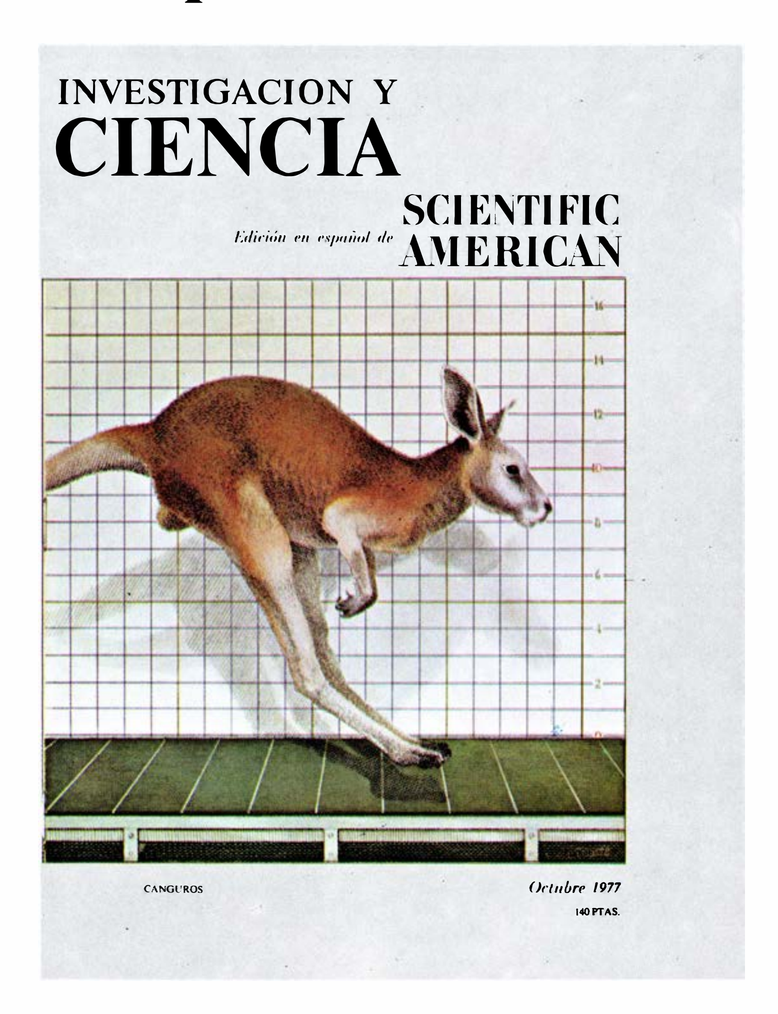
GENERALIZED "SANDWICH" THEOREMS. A. H. Stone y J. W. Tukey en *Duke Mathematical Journal*, vol. 9, n.º 2, págs. 356-359; junio, 1942.

TALLER Y LABORATORIO

THE RAINBOW: FROM MYTH TO MATHEMATICS. Carl B. Boyer. Thomas Yoseloff, 1959.

MULTIPLE RAINBOWS FROM SINGLE DROPS OF WATER AND OTHER LIQUIDS. Jearl D. Walker en *American Journal of Physics*, vol. 44, n.º 5. págs. 421-433; mayo, 1976.

Seguiremos explorando los campos del conocimiento



LAS NEGOCIACIONES SALT, Herbert Scoville, Jr.
Las conversaciones en curso sobre la limitación de armamentos estratégicos pueden poner un techo eficaz a la carrera cuantitativa armamentística y constreñir la carrera cualitativa. El equilibrio actual podría romperse si esto no se consigue.

OBJETOS BL LACERTAE, Michael J. Disney y Philippe Véron
Se cuentan entre los objetos del universo más alejados y más luminosos. La fuente de su radiación se encuentra en un pequeño núcleo (cuyo tamaño pudiera ser el del sistema solar) que despide más luz que una galaxia.

COMUNICACIONES POR ONDAS DE LUZ, W. S. Boyle
El primer ensayo comercial con servicios telefónicos por ondas luminosas se encuentra ya funcionando en Chicago. Las señales se transmiten a través de fibras de cristal en pulsos generados por finas fuentes luminosas.

FLUJO DEL CALOR DESDE EL INTERIOR DE LA TIERRA, Henry N. Pollack y David S. Chapman
Se puede confeccionar un mapa sobre el flujo del calor global a partir de las mediciones de miles de campos en continentes y en suelos oceánicos. El modelo resultante se interpreta en términos de la tectónica de placas.

CANGUROS, T. J. Dawson
Las estrategias de adaptación seguidas por estos mamíferos son afines y paralelas a las asumidas por los ungulados de las sabanas semiáridas del Viejo Continente.

LOS VACUOLOS GASIFEROS DE LAS CIANOFICEAS, A.E. Walsby
Las células de muchas especies acuáticas de estos organismos parecidos a bacterias contienen miles de estructuras diminutas llenas de gas.

PERSPECTIVAS EN QUIMICA MACROMOLECULAR, Juan A. Subirana
La hipótesis macromolecular ha permitido tanto el desarrollo de la gran industria de los plásticos como penetrar en los secretos de la vida.

SEGUNDOS MENSAJEROS EN EL CEREBRO, James A. Nathanson y Paul Greengard
Las células nerviosas se comunican entre sí por medio de la secreción de neurotransmisores. Estas substancias constituyen mensajes químicos que, a través de unos segundos mensajeros que se hallan en el interior de la célula, son traducidos a una serie de efectos fisiológicos que pueden ser pasajeros o de larga duración.

INVESTIGACION Y
CIENCIA